NOVEL NONGENETIC ALLELE-SPECIFIC EXPRESSION

IN THE MAMMALIAN BRAIN

by

Wei-Chao Huang

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Neurobiology and Anatomy

The University of Utah

December 2017

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Wei-Chao Huang**

has been approved by the following supervisory committee members:

Christopher Gregg	, Chair	**9/15/2017** Date Approved
Bradley Cairns	, Member	**9/15/2017** Date Approved
Jan Christian	, Member	**9/15/2017** Date Approved
Adam Douglass	, Member	**9/15/2017** Date Approved
Charles Murtaugh	, Member	**9/15/2017** Date Approved

and by **Monica Vetter** , Chair/Dean of

the Department/College/School of **Neurobiology and Anatomy**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Most genes in a diploid organism are thought to be expressed and regulated equally by the two parental alleles. However, a subset of genes referred to as monoallelically-expressed genes are expressed preferentially only on the maternal or paternal allele. Monoallelically-expressed genes include imprinted genes, which express either the maternal or paternal allele preferentially, as well as random monoallelic genes that choose to express one allele at random. Examples of in vivo random monoallelic genes include X-linked genes in females, immunoglobulins in B cells, clustered protocadherins, and olfactory receptors. However, these examples generally are thought to be rare in the mammalian genome and confined to a few genes that have a uniquely clustered organization in the genome. In this dissertation, new genome-wide analysis methods were developed to reveal the landscape of epigenetic allelic effects in the mouse. Chapter 2 details the discovery of novel, noncanonical genomic imprinting effects that are enriched uniquely in the brain. Chapter 3 details the development of new genomics and statistical methods to identify genes that express their alleles differentially in vivo through nongenetic mechanisms. This novel screening strategy was applied to reveal the landscape of epigenetic allelic effects in the mouse brain at different ages and to profile different tissue types. The dissertation describes the discovery of diverse forms of nongenetic differential allelic expression in mouse, macaque, and human brains, and demonstrates that these effects interact with heterozygous mutations and shape genetic

architecture at the allelic and cellular level in vivo. The results challenge current thinking in the fields of genetics and epigenetic by revealing a new landscape of epigenetic allelic effects in vivo, and describe new mechanisms that may shape phenotypic variation and susceptibility to disease. Although the functions of these allelic effects are unknown, and the regulatory mechanisms involved have yet to be defined, the future of the field is discussed.

Keywords: Allele-specific expression, Genomic imprinting, Imprinted genes, Monoallelic expression, Epigenetic, Random monoallelic expression, X chromosome, X-inactivation, RNA-seq, RNA sequencing, Single cell sequencing, Autism

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Working with Dr. Christopher Gregg, I had the freedom to explore ideas that may have initially seemed absurd. He always patiently organized my thoughts, gently corrected my mistakes, and guided me to turn those crazy ideas into high-quality research. I'll be forever thankful for all Chris has done for me. I also would like to thank present and past members of Gregg lab for their great help. Drs. Paul Bonthius and Cornelia Hörndli are like my older brother and sister and assisted me with everything within and outside the lab. Elliott Ferris is my diviner to interpret every experimental result I couldn't understand. He also helped me keep my Mandarin from getting rusty. Tina Cheng and Susan Steinwand conscientiously organized Gregg lab and supported every lab member through many experiments. I am thrilled to be working with Kathleen Bennett and Stephanie Kravitz, and I believe they will lead the Gregg lab to a bright future.

I would like to thank the members of my dissertation committee Drs. Bradley Cairns, Jan Christian, Adam Douglass, and Charles Murtaugh for their support, guidance, and encouragement. I feel fortunate to be a part of the Neurobiology and Anatomy Department at the University of Utah, which has provided a scientifically diverse environment to expand my vision and make me a better scientist. I am grateful to Dr. Sheryl Scott and Ms. Nicole Caldwell for their thoughtful reminders to help me fulfill my duties of a PhD student. I'd like to thank the friends who continually support me,

especially Dr. Chun-Han Lin at UC Berkeley, Yixuan Guo, and Yueqi Wang at the University of Utah and several friends in Taiwan and the U.S.

Finally, I would like to thank my beloved family including my parents, my younger brother, and my aunts and their families. My parents have always provided endless love to me. They understood the challenges I faced and were supportive of my decisions to stay oversea. I am also extremely grateful for everything my younger brother Wei-jen has done. He took care of my parents -especially when my father was in the hospital due to aortic dissection last year. Without him, I would not have been able to stay here to finish my PhD.

CHAPTER 1

INTRODUCTION

Part of this chapter will be incorporated in a review in preparation Gregg C & Huang WC *Trends in Neurosciences*, 2018.

1.1 Overview

In diploid organisms, genes are assumed to regulate and express maternal and paternal alleles equally. However, the discovery of genomic imprinting in the 1970s and 1980s challenged this assumption and demonstrated that at least some genes express their maternal and paternal alleles differently because of epigenetic effects. The realization that this can occur transformed thinking in the genetics field at the time; however, at the time, very few genes were thought to be affected. Subsequently, the field has sought to understand allele-specific expression effects more thoroughly. This section provides an overview of the discovery, functions, and mechanisms of allele-specific epigenetic effects.

1.2 Monoallelically-expressed genes

Genes with allele-specific expression (ASE), also described as monoallelic expression, can be classified predominantly as two distinct groups (Chess, 2013, 2016). The first is deterministic and includes genomic imprinting, in which genes express one parental allele preferentially over the other. The second involves random monoallelic expression (RME), in which the maternal or paternal allele is selected for expression randomly. X-linked genes in females are a group of genes that has been studied well and shows RME attributable to random X-inactivation for dosage compensation (Nguyen & Disteche, 2006). However, some autosomal genes are known well to exhibit RME in vivo, including antigen receptors, clustered protocadherins, and olfactory receptors, and the state of knowledge about these different forms of epigenetic allelic effects is discussed below.

1.3 Genomic imprinting

1.3.1 Parent-of-origin effects

1.3.1.1 Parent-of-origin effects in mouse and human

The discovery of genes with genomic imprinting arose from genetic studies of parent-of-origin effects on phenotypes. Johnson (1974) published the first genetic study of mice that suggested that parent-of-origin effects influence morphology and lethality in mammalian offspring. The author found that heterozygous offspring that inherit a maternal mutation on chromosome 17, called the hairpin-tail allele, have smaller litter sizes and strong morphological defects in the tail. In contrast, offspring that inherit the paternal mutant allele are viable and have reduced phenotypic effects. The hairpin-tail allele was later shown to contain an imprinted gene expressed maternally, called the insulin growth factor 2 receptor (*Igf2r*; Barlow, Stoger, Herrmann, Saito, & Schweifer, 1991). In other studies of parent-of-origin effects that used Robertsonian translocations that can cause duplicated or deficient chromosomal regions, Cattanach and Kirk produced chromosomally-imbalanced mice that carried a pair of homologous chromosomes derived from only one parent to obtain uniparental disomy (Cattanach & Kirk, 1985). They found that the mice with uniparental disomy of chromosome*s* 1, 3, 4, 5, 9, 13, 14, or 15 survived, while maternal disomy on chromosome 6 was lethal*,* as were regions of chromosomes 2, 8, and 17 with maternal disomy. At approximately the same time, Searle and Beechey performed reciprocal translocations to produce mice with uniparental disomy, and their findings were consistent with those of Cattanach and Kirk (Searle & Beechey, 1978). Those studies identified several subchromosomal regions that require both maternal and paternal inheritance for normal embryonic development*,* and provided

important support for subsequent studies of the maternal and paternal complements of the genome using pronuclear transplantation, as detailed below.

The first human study to report parent-of-origin effects examined a pedigree with familial Beckwith-Wiedemann Syndrome (BWS). Lubinsky and colleagues found that offspring affected with BWS were born only to female and not male carriers (Lubinsky, Herrmann, Kosseff, & Opitz, 1974). Therefore, they postulated that the inheritance must involve unique factors mediated by the ovum but not the sperm. We now know that a subset of BWS is caused by mutations or epigenetic dysregulations of the cyclin-dependent kinase inhibitor 1C (*CDKN1C*) gene, a cell cycle regulator imprinted maternally (Maher & Reik, 2000).

Studies of uniparental disomy in humans provided further evidence of parent-of-origin effects. A disorder caused by uniparental disomy was first identified when a female child was diagnosed with cystic fibrosis and short stature attributable to a duplication of maternal chromosome 7 that lacked the paternal copy (Spence et al., 1988). These symptoms of short stature and body asymmetry are now referred to as Silver–Russell syndrome, which is a genetic disorder caused by mutations that affect the *H19/IGF2* imprinted gene locus. Overall, these early studies of parent-of-origin effects in mice and humans laid the foundations for the discovery of genomic imprinting.

1.3.1.2 Pronuclear transplantation

Scholars agree that the decisive experiments that proved the existence of genomic imprinting were those Solter and Surani (McGrath & Solter, 1983; Surani & Barton, 1983) performed on pronuclear transplantation. This technique is based on the fusion of a

pronucleus with enucleated cytoplasts from a newly-fertilized egg. For example, a newly-fertilized egg is removed and the paternal pronucleus is replaced with a maternal pronucleus to generate a bimaternal diploid embryo (Gynogenote). Similarly, the maternal pronucleus could be replaced with another paternal pronucleus to engineer an embryo with two paternal genomes (Androgenote). The results of these engineering experiments showed that both gynogenetic and androgenetic embryos failed to survive, while those engineered with one maternal and one paternal pronucleus were viable (McGrath & Solter, 1984; Surani, Barton, & Norris, 1984).

Remarkably, gynogenetic and androgenetic embryos exhibited recipocal phenotypes of developmental failure: defects in extraembryonic tissues caused the death of androgenetic embryos, while the death of gynogenetic embryos was attributable to the failed development of embryonic tissue (McGrath & Solter, 1984; Surani et al., 1984). These results suggested that the maternal genome is necessary for the growth of the embryonic lineage, while the paternal genome plays an essential role in extraembryonic development.

To study later stages of development, subsequence studies engineered chimeric mice by mixing gynogenetic or androgenetic cells with normal cells. These chimeric embryos survived longer, but exhibited unique phenotypes. The studies showed that gynogenetic cells contributed preferentially to primitive ectoderm lineages (including yolk sac mesoderm), while the androgenetic component was restricted to trophoblast and primitive endoderm lineages (Barton, Ferguson-Smith, Fundele, & Surani, 1991; Fundele et al., 1990). These pronuclear transplantation studies provided strong evidence of genomic imprinting by revealing that maternal and paternal chromosomes are not

equivalent.

1.3.2 Identification of imprinted genes

The first three imprinted genes in mice were described in 1991 by different research groups. Barlow et al. demonstrated that the *Igf2r* gene expresses the maternal copy alone, and causes the maternally-derived phenotype in the Hairpin-tail mouse (Barlow et al., 1991). A few months later, DeChiara, Robertson, and Efstratiadis (DeChiara, Robertson, & Efstratiadis, 1991) identified insulin growth factor 2 (*Igf2*) on chromosome 7, and showed that *Igf2* expresses only the paternal allele and represses the expression of the maternal allele. Ferguson-Smith, Cattanach, Barton, Beechey, and Surani (Ferguson-Smith, Cattanach, Barton, Beechey, & Surani, 1991) also showed that *Igf2* expression is repressed in embryos with maternal uniparental duplication and deficiency of the paternal chromosome harboring this gene, revealing preferential expression of the paternal allele. Bartolomei, Zemel, and Tilghman (Bartolomei, Zemel, & Tilghman, 1991) mapped the third gene, *H19*, to distal chromosome 7 and showed that it was an imprinted gene expressed maternally. Diverse methods were used to identify the first three imprinted genes. Positional cloning was used to map *Igf2r* in the Hairpin-tail deletion in chromosome 17. For *Igf2*, the physiological function and expression pattern was tested by gene knockout technology combined with in-situ hybridization, and *H19* was identified by the RNase protection assay in hybrid mice tissues derived from reciprocal crosses of different strains.

The discovery that *H19* on chromosome 7 is spatially close to *Igf2* led to a hypodissertation that imprinted genes could be clustered together, and indeed, many more

imprinted gene loci were found subsequently. For example, *Igf2r/Airn* cluster (Wutz et al., 1997), *Nespas/Gnas* cluster (Peters et al., 1999), *Snrpn/Ube3a* cluster (Buiting et al., 1995), and so on, were identified during the 10 years after the first imprinted genes were discovered, and to date, at least 150 imprinted genes and ~15 imprinted clusters have been identified in the mouse genome (http://www.geneimprint.com/). The clustering of imprinted genes suggests that the mechanisms of imprinting control could work in a cis-acting manner. The regulatory mechanisms of genomic imprinting are reviewed below.

1.3.3 Mechanisms of genomic imprinting

1.3.3.1 Nongenetic regulatory mechanism and CG repeat

The initial observation that more than 80% of imprinted genes are located in gene clusters (Wan & Bartolomei, 2008), and not scattered in the genome, suggested that the mechanisms of genomic imprinting could be cis-acting, and involve regulatory elements that control gene expression uniquely on one chromosome (Reik & Walter, 2001). The DNA sequence of homologous chromosomes typically is identical except for some random polymorphic sites that differ among individuals. This indicates that the main factors that control genomic imprinting are epigenetic processes that mark the DNA sequence to activate or repress one chromosome relative to another, rather than because of the alteration of the DNA sequence between two chromosomes. Subsequent studies were designed to identify the shared sequence in known imprinted clusters to predict the presence of imprinted loci/genes (Luedi, Hartemink, & Jirtle, 2005). The predictive shared sequence accelerates the finding that regulatory elements underlie genomic imprinting. The CG-rich repeat was found to be enriched in several imprinted clusters

(Hutter, Helms, & Paulsen, 2006; Paulsen et al., 2000), and the repeats correspond to the regions subject to mono-parental methylation (Smilinich et al., 1999).

1.3.3.2 DNA methylation

DNA methylation is an epigenetic modification in which a methyl group is added to a cytosine or adenine/adenosine base (Chen, Zhao, & He, 2016; Ratel, Ravanat, Berger, & Wion, 2006). The CpG island is the region enriched for repeat CpG dinucleotides**,** and plays a major role in establishing DNA methylation (Bird, 1986). Methylation of cytosine (C) in the CpG island is associated typically with transcriptional repression, especially in the promoter region of a gene, while demethylation is associated with activation of gene expression. 60-80% of DNA methylation in vertebrates is located in a CpG context in somatic cells (Meehan, Lewis, McKay, Kleiner, & Bird, 1989), and dysregulation of DNA methylation has been implicated in various human disorders (Bergman & Cedar, 2013). The brain has higher levels of methylated cytosines compared to other organs**,** especially in repetitive sequences, and long-term memory consolidation may be influenced by DNA methylation (J. J. Day & Sweatt, 2010; Miller & Sweatt, 2007). The common feature in imprinted clusters is that each contains at least one differential methylation region (DMR), which is a genomic region in which the maternal and paternal alleles are methylated differentially (Barlow & Bartolomei, 2014).

The mechanisms that mark two parental chromosomes differentially are predicted to involve particular features (Ferguson-Smith, 2011). First, the individual parental chromosomes should be modified before fertilization, when they are not in the same nucleus. Second, the modification must influence gene transcription. Third, it must be

maintained and inherited during DNA replication after fertilization. Finally, the modification must be reversible and erasable in the germline when the establishment of a new marker is needed according to the sex of the offspring. It was postulated that DNA methylation is the optimal epigenetic modification that fulfills all four features.

When is DNA methylation of imprints established? During mouse development, the imprinting life cycle has three phases: *erasure, establishment,* and *maintenance*. Erasure of imprinting begins when primordial germ cells (PGCs), the common origin of the germline, migrate to the genital ridge. The first wave of global demethylation is observed in PGCs between E7.25 to E9.5; however, the methylation of imprinting control regions is maintained during this phase (McLaren, 2003; Seisenberger et al., 2012). Instead, imprinting control regions are demethylated during a subsequent event between E10.5 to E13.5, when PGCs enter the genital ridge. At this time, imprinted genes are expressed or silenced biallelically (Seisenberger et al., 2012). The erasure of DNA methylation is thought to be accomplished by two broad mechanisms: passive and active demethylation, although the details are not yet understood fully. Passive demethylation occurs in the absence of methylation of newly-synthesized DNA strands by inactivating the methylation machinery during DNA replication. The active removal of DNA methylation take places via direct conversion of a 5-methylcytosine to cytosine. One of the major enzymes that has active DNA demethylation capability is the ten-eleven translocation 1 (TET) enzyme. TET1 can catalyze oxidation of 5-methylcytosine to produce 5- hydroxymethylcytosine (5hmC), which is thought to be the first step in DNA demethylation. Thereafter, TET enzymes oxidize 5hmC further to 5-formylcytosine and 5-carboxylcytosine, which are the intermediates in the process of converting 5hmC back

to cytosine. The work by Yamaguchi, Shen, Liu, Sendler,and Zhang (Yamaguchi, Shen, Liu, Sendler, & Zhang, 2013) recent work on *Tet1*-deficient mice indicated that the TET1 enzyme is important in the demethylation of imprinted DMRs. *Tet1*-deficient males showed aberrant methylation patterns at imprinted loci in PGCs and sperm cells, and the offspring produced by Tet1-deficient male mice crossed with wild-type females exhibited a number of phenotypes associated with abnormal imprinting erasure that indicated strikingly that *Tet1* may mediate parental imprinting erasure during PGCs' development.

The erasure of methylation marks occurs between E6.5 and E9.5, before PGCs initiate differentiation. While differentiation occurs, the new imprint can be acquired in a sex-specific manner in the germline. Interestingly, the establishment of new paternal and maternal imprints in germ cells occurs at different times (Davis, Trasler, Moss, Yang, & Bartolomei, 1999; J. Y. Li, Lees-Murdock, Xu, & Walsh, 2004). De novo DNA methylation in female germ cells occurs in the mature oocytes *postnatally* when the growing oocytes enter a transcriptionally silent stage. In contrast, paternal imprints take place *prenatally* during mitotic arrest of spermatogonia precursor cells (Smallwood & Kelsey, 2012). The DNA methyltransferase 3 (DNMT3) family regulates de novo methylation (Kaneda et al., 2004). The de novo methyltransferase gene *Dnmt3a* is required to establish imprints in both male and female germ line cells. In addition, *Dnmt3l* is a modulator that has no methyltransferase activity, but has been shown to cooperate with *Dnmt3a* and *Dnmt3b* to establish imprints in mouse oocytes (Bourc'his, Xu, Lin, Bollman, & Bestor, 2001; Jia, Jurkowska, Zhang, Jeltsch, & Cheng, 2007; Suetake, Shinozaki, Miyagawa, Takeshima, & Tajima, 2004). DNMT3A and DNMT3B function in a complex with DNMT3L that works with histone deacetylase and

methyltransferases for histone 3 lysine 9 (H3K9), while DNMT3L recognizes nucleosomes that lack histone 3 lysine 4 (H3K4) methylation selectively (Ooi et al., 2007).

The final phase of the imprinting life cycle is maintenance. DNA methylation at imprinting control regions must be maintained following cellular mitosis. DNA methyltransferase 1 (*Dnmt1*) has been found to have a high affinity for nonmethylated CpG islands in a biochemical assay (Hermann, Goyal, & Jeltsch, 2004; Yoder, Soman, Verdine, & Bestor, 1997). It recognizes unmethylated CpGs on one of two strands, which is also described as hemimethylated DNA. To achieve maintenance of DNA methylation, DNMT1 restores to symmetric methylation the progeny DNA strand from a state that contains one methylated parental stand and one newly-replicated strand caused by semiconservative DNA replication (Vilkaitis, Suetake, Klimasauskas, & Tajima, 2005). Indeed, Li, Beard, and Jaenisch's (E. Li, Beard, & Jaenisch, 1993) seminal study on *Dnmt1*-deleted mouse proved that *Dnmt1* is required for genomic imprinting. Another important maintenance of imprints is mediated during the early embryonic stage. A second wave of genome-wide demethylation occurs during early embryogenesis, by which unnecessary germ-line specific methylation is removed, while marks associated with imprinted genes are preserved, enabling parent-of-origin effects in later tissues. Several factors have been proposed to be involved in the protection of imprinting from demethylation in the early embryo. *Zfp57* null mice showed derepression of the imprinted allele in the imprinted clusters and embryonic lethality (X. Li et al., 2008). Maternal deletion of *Trim28* also leads to embryonic lethality attributable to misregulation of genomic imprinting (Messerschmidt et al., 2012). Interestingly, Quenneville and

colleagues' (Quenneville et al., 2011) study showed that ZFP57 that binds with TRIM28, a component of a multifunctional repressor complex consisting of H3K9 (H3K9me3)-catalyzing histone methyltransferase SETDB1, the heterochromatin protein 1 (HP1), DNMT1, DNMT3A, and DNMT3B, is detected at the imprinted allele in DMRs in embryos.

1.3.3.3 Histone modifications

Covalent posttranslational modifications to histone proteins, which include methylation, phosphorylation, acetylation, and ubiquitylation, function to influence the overall structure of chromatin and the binding of effector molecules to modulate gene expression. The combinatorial nature of histone modifications has been proposed to modulate the transcription of genetic information (Jenuwein & Allis, 2001). Allfrey, Faulkner, and Mirsky (Allfrey, Faulkner, & Mirsky, 1964) first identified histone modification in 1964; however, the entire landscape of modification remains elusive. Acetylation and methylation of lysine are two modifications that have been studied well, and the enzymes that add or remove these two markers on the histone have been identified. For lysine acetylation, histone acetyltransferases (HATs) transfer an acetyl group to lysine, while histone deacetylases (HDACs) reverse lysine acetylation. Histone lysine methyltransferase (HKMT) deposits methyl groups on lysine, and SUV39H1, which regulates di/trimethylation of H3K9, was the first enzyme found to mediate this process (Rea et al., 2000). Subsequent studies (Cao et al., 2002; L. Wang et al., 2004) showed that the polycomb repressive complex 2 (PRC2) mainly ensures di/trimethylation of H3K27. Histone methylation previously was considered stable and static, but several

lysine demethylases were characterized following the first lysine demethylase LSD1 identified (Shi et al., 2004). Generally, regions of acetylated histones are associated consistently with gene activity, while the functions of histone methylation are more site-specific. For example, H3K4, H3K36, and H3K79 methylation are found on activated genes, while H3K9 or H3K27 methylation is associated with repressed genes (Jenuwein & Allis, 2001; Sims, Nishioka, & Reinberg, 2003).

Because of their roles in the stable repression of gene expression, H3K27 and H3K9 are good candidates for roles in repression of the silent allele for imprinted genes; on the other hand, H3K4 or H3K79 may mark on the expressed allele. Xin Allis, and Wagstaff (Xin, Allis, & Wagstaff, 2001) observed that H3K9 is methylated on the silent maternal copy of the *SNRPN/UBE3A* imprinted locus, while H3K4 is methylated on the expressed paternal copy. Other imprinted regions also have been found to follow a similar pattern of allele-specific histone modifications (Fournier et al., 2002). In hybrid mice crosses of two different strains, diverse histone modifications have been examined in an allele-specific manner, and have revealed a reciprocal pattern of activation and repression of allele-specific histone marks at imprinting control regions methylated differentially.

Genetic studies of mice also have provided evidence that indicates that histone modifications play a role in genomic imprinting. Magnuson, Montgomery, de Villena, and Magnuson (Mager, Montgomery, de Villena, & Magnuson, 2003) examined the imprinting pattern in embryos with a deficient embryonic ectoderm development (*Eed*) gene, a main component in PRC2, and found that a subset of imprinted genes, particularly *Ascl2*, *Grb10*, *Cdkn1c*, and *Gtl2*, turned to biallelical expression in mutant

embryos. H3K9 methylase *G9a* is also implicated in the regulation of imprinted genes. Genetic ablation of *G9a* in the placenta shifted a maternal imprinted gene, *Slc22a3*, to a biallelic expression (Nagano et al., 2008), and mouse embryonic stem cells lacking *G9a* exhibited loss of DNA methylation in DMRs of several imprinted genes (T. Zhang et al., 2016). Kim and colleagues' (Y. Kim et al., 2017) recent study showed that chemical interference with *G9a* can activate the silent copy of certain imprinted genes and improves survival and growth in a mouse model of Prader–Willi syndrome, a disorder caused by deletion of the paternal copies of the *SNRPN/UBE3A* imprinted cluster.

Histone modifications in various genomic elements have been shown to be associated with gene activity and repression, which indicates a functional relation between DNA methylation and histone modifications. As mentioned above, DNMT3L, a factor required to reestablish genomic imprinting in germ cells, recognizes nucleosomes that lack H3K4 methylation selectively (Ooi et al., 2007). Ciccone (Ciccone et al., 2009) demonstrated that females deficient in KDM1B, a H3K4 demethylase enzyme, exhibited increased H3K4 methylation and unmethylated DMRs in four imprinted regions in the *Kdm1b* null oocyte. A subsequent study (Guo et al., 2015) proposed a biochemical mechanism in which the unmethylated histone H3K4 tail is able to stimulate the enzymatic activity of DNMT3A through a conformational change of DNMT3A from an autoinhibited to an active form.

1.3.3.3 Noncoding RNA (ncRNA)

Advancement in next-generation sequencing technologies has allowed large-scale RNA profiling and revealed that >75% of the human genome is a highly complex

network of messenger RNA (mRNA) and ncRNA (Djebali et al., 2012). Long noncoding RNAs (lncRNAs), usually longer than 200 nucleotides and with no apparent protein-coding role, are found in each branch of life, and have been implicated in highly-specialized regulatory functions (Rinn & Chang, 2012). lncRNAs are diverse and numerous; according to a recent review (H. Wu, Yang, & Chen, 2017), at least nine categories of lncRNAs have been characterized in the eukaryotic transcription system, and the number of lncRNAs exceeds the number of protein-coding genes.

Several functions of lncRNAs have been proposed in recent years; however, lncRNAs' role in the regulation of genomic imprinting was demonstrated a relatively long time ago. One example that showed the lncRNA-mediated imprinting model is the *Igf2r* cluster (Figure 1.1), in which lncRNA is expressed paternally, and *Airn*, which is antisense to *Igf2r* in the gene cluster (Sleutels, Zwart, & Barlow, 2002; Wutz et al., 1997). The *Igf2r* imprinted gene cluster contains several protein-coding genes (*Igf2r*, *Slc22a2*, *Slc22a3*) expressed maternally, and the single DMR corresponding to the maternal imprinting control region in this cluster is located between exons 2 and 3 of the *Igf2r* gene (Stoger et al., 1993). The *Airn* promoter is located within the imprinting control regions that are subject to differential methylation, and is hypermethylated on the maternal allele. By interacting with repressive histone modifications, the hypermethylated promoter interferes with the transcription of *Airn* on the maternal allele, while on the other hand, binding the paternal unmethylated region with the CCCTC-binding factor (CTCF) insulator results in the expression of *Airn* on the paternal allele. Thus, the *Airn* expressed paternally represses *Igf2r* expression on the paternal allele by transcriptional interference. Truncation of the *Airn* transcript generated by insertion of a

polyadenylation site results in the biallelic expression of *Igf2r* (Sleutels et al., 2002), indicating that full-length *Airn* transcriptional interference is required to silence the paternal Ig2r allele for genomic imprinting. Here, *Airn* in the *Igf2r* cluster exemplifies the model of lncRNA-mediated imprinting; other imprinted loci mediated by lncRNAs include the *Kcnq1* locus, *Gnas–Nespas* gene cluster, and *Igf2/H19* locus (Santoro & Barlow, 2011).

1.3.3.4 Insulator

Insulators are DNA-binding proteins that hamper the interaction between enhancers and promoters. The CCCTC-binding factor (CTCF), also known as the 11-zinc finger protein, is an essential insulator of interactions between regulatory elements. CTCF was characterized initially as a negative regulator in the chicken *CMYC* gene, and its name reflects the finding that it binds the CTC-rich sequence of the *CMYC* gene (Filippova et al., 1996; Lobanenkov, Nicolas, Plumb, Wright, & Goodwin, 1986). However, subsequent studies have revealed that the CCCTC motif alone is neither necessary nor sufficient for CTCF protein binding to DNA (Chernukhin et al., 2000). Using genome-wide mapping, several studies have shown the widespread occupancy of CTCF in the genome. For example, Wang et al. found that CTCF binds to an average of approximately 55,000 DNA sites from 19 diverse cell types in vitro, while in an in vivo study, Prickett et al. (Prickett et al., 2013) also identified ~50000 CTCF-bound sites, which is consistent with the hypodissertation that CTCF plays a role in whole-genome chromatin organization.

The mechanisms involved in CTCF's regulation of gene transcription also have

been explored. By binding specific DNA sequences, CTCF can block promoter-enhancer interactions to repress gene expression (Hark et al., 2000). CTCF binds to DNA in a methylation-sensitive manner, which makes it suited well to respond to DMRs and regulate imprinting (Hark et al., 2000). In addition, CTCF is able to bind to itself to form homodimers that mediate the formation of DNA loops, and plays a role in the regulation of nuclear architecture to induce chromosomal loops that regulate gene expression over long distances in the genome (Splinter et al., 2006). Cohesin, a multisubunit protein complex comprised of four subunits, interacts with CTCF throughout the genome and is essential to establish DNA loops and insulation (B. K. Lee & Iyer, 2012). The CTCF/cohesin complex has been shown to mediate genomic imprinting in a subset of imprinted clusters (Franco, Prickett, & Oakey, 2014).

Here, the *Igf2/H19* imprinted cluster is used as an example to illustrate the way in which CTCF and chromosomal looping contribute to imprinting (Figure 1.2). In the *Igf2/H19* imprinted cluster, the protein-coding gene, *Igf2*, is expressed from the paternal allele and *H19* is expressed from the maternal allele. This reciprocal expression pattern depends on one DMR corresponding to a paternally-imprinted control region, which is located in the intergenic region between *Igf2* and *H19*. The CTCF/cohesin complex is able to bind to the unmethylated maternal allele, while its binding is blocked on the methylated paternal allele because of the methylation-sensitive feature (Pant et al., 2004). *Igf2* and a downstream noncoding RNA gene*, H19,* share multiple enhancers that are located close to the 3' end of H19. Chromatin conformation capture showed long-range interactions between enhancers and the promoter of the *Igf2* gene for the paternal allele, which contains the methylated DMR that inhibits the occupancy of CTCF (Kurukuti et

al., 2006; Murrell, Heeson, & Reik, 2004; Yoon et al., 2007). In contrast, the unmethylated maternal allele can bind with the CTCF/Cohesin complex, which forms a chromatin loop to hinder the enhancer interaction with *Igf2* spatially, and leads to redirection of the enhancer to the *H19* promoter. Thus, *Igf2* and *H19* compete for access to the same upstream enhancers, and CTCF that binds to one allele in the differentially-methylated imprinting control region regulates the formation of distinct chromatin loops on each allele.

1.3.4 Biological significance of genomic imprinting

After the first imprinted gene was identified, studies that sought to identify the functions of imprinted genes increased rapidly. The next section presents an overview of three aspects of the biological significance of genomic imprinting: evolutionary origin, human imprinting disorders, and imprinted gene functions in the brain.

1.3.4.1 Evolutionary origin of genomic imprinting

Several theories have been proposed that attempt to address fitness advantages associated with genomic imprinting that explain the way in which it evolved. Largely, these theories postulate that the cooperation or conflict between two parental alleles expressed differentially is the core feature that leads to selection for genomic imprinting. The kinship and coadaptation theories are the leading theories. The kinship theory is a kin selection model Haig and Westoby (1989) proposed based on Trivers' concept that parent investment is "…any investment by the parent in an individual offspring that increases the offspring's chances of surviving at the cost of the parent's ability to invest in other

offspring" (Haig, 2014). In promiscuous mating systems, an individual father frequently is related only to a subset of the offspring, while the mother nurtures all of the offspring equally. This difference results in contrasting evolutionary interests on the part of the parental genome, such that, for the father to achieve optimal fitness, it is evolutionarily advantageous for the paternal genes to increase the demand to acquire maximum maternal resources that are invested in the fetus or infant so that the offspring has a greater likelihood of future survival. On the other hand, it is evolutionarily advantageous for maternal genes to reduce the demand for maternal energy, conserve it for future pregnancies, and distribute their resources equally to all progeny. Hence, the kinship theory predicts that the function of genes expressed paternally is to increase offspring demand for maternal resources to enhance fetal growth, while those expressed maternally decrease demand to suppress growth (Haig & Graham, 1991; Wilkins & Haig, 2003). Some experimental findings support this theory; for example, *Igf2*, which is expressed paternally, is a growth factor that increases offspring size, but the *Igf2r*, expressed maternally, is a growth suppressor that reduces offspring size.

On the other hand, the coadaptation theory postulates that imprinted genes provide phenotypic compatibility between mother and offspring to optimize fetal development (Wolf & Hager, 2006). This theory predicts that the genetic interaction between offspring and mother increases fitness when the offspring's genome expresses only a maternal allele. In contrast, the maternal allele tends to be silenced, while the paternal allele is expressed if a fitness interaction is achieved best by mismatched coexpression. Some evidence supports this theory, such as *Grb10*, which is expressed maternally (Cowley et al., 2014), or *Peg3*, a gene that is expressed paternally (Curley,

Barton, Surani, & Keverne, 2004).

The interaction between mother and offspring as an evolutionary battleground (kinship theory) or as a coordinated effort with a common goal (coadaptation theory) is still debated; however, they are not necessarily exclusive. Haig (2014) attempted to reconcile the two theories, and stated, "cooperation and conflict are two sides of one coin." However, both the kinship and coadaptation theories have been challenged by some contradictory examples. Other evolutionary hypotheses about genomic imprinting have been proposed, for example, the sexual antagonism theory (T. Day & Bonduriansky, 2004) or the nonconflict theory (Spencer & Clark, 2014). However, their ability to explain actual imprinted genes has yet to be tested.

1.3.4.2 Human imprinting disorders

The first imprinting genetic disorder described in humans is Prader-Willi Syndrome (PWS), which is characterized by early childhood obesity, poor growth and physical development, cognitive impairment, and behavioral problems (Nicholls, Knoll, Butler, Karam, & Lalande, 1989). In fact, PWS generally has two phases in clinical development. In infants, PWS is characterized by poor muscle tone, and diminished swallowing and suckling. After the age of two, the features of PWS are insatiable appetite and food-seeking behavior, developmental delay, and psychomotor retardation. Angelman syndrome (AS), which exhibits different clinical symptoms, and is characterized by speech impairment, abnormal happiness with unprovoked laughter, severe motor and neurologic deficits, epilepsy, and seizures, is thought to be PWS's sister disorder because both disorders result from the deletion or mutation of the paternal or

maternal contribution to the same region of chromosome 15q11-13 (Cassidy, Dykens, & Williams, 2000). The 15q11-13 imprinted cluster includes a protein coding gene, *UBE3A*, expressed maternally, and several noncoding RNAs expressed paternally. PWS is caused by the loss of paternal expression of the 15q11-13 region, which includes a cluster of small nucleolar RNAs, while AS results from the loss of *UBE3A* in the 15q11-13 region.

To date, at least 12 human imprinting disorders have been identified at diverse imprinted loci in chromosome 2, 6, 7, 8, 9, 14, 15, and 20 (Peters, 2014; Soellner et al., 2017). The common clinical manifestations of these imprinting disorders include: (1) abnormal pre/postnatal growth, (2) dysregulation of blood sugar, (3) aberrant food-seeking behavior, (4) mental retardation with behavior problems, and (5) precocious puberty (Soellner et al., 2017). However, these shared clinical features make it difficult to distinguish between the different disorders to reach an accurate diagnosis. In addition, conducting molecular testing for an imprinting disorder is challenging. Several genetic/epigenetic factors have been found to cause imprinting disturbances, such as copy number variation, uniparental disomy, aberrant methylation pattern, and point mutation in imprinted genes, and increasing evidence indicates that genetic/epigenetic changes could occur in multiple imprinted loci, which confounds the identification of imprinting disorders (Sanchez-Delgado et al., 2016). Further, the mosaicism of imprinting disturbances in different tissues hampers accurate diagnosis. Thus, the development of a comprehensive and precise diagnostic assay for imprinting disorders is still needed.

Early experiments that involved the generation of mouse chimeras by mixing parthenogenetic cells (PC) or androgenetic cells (AC) with normal cells suggested that imprinted genes may play important roles in the brain (Allen et al., 1995). The authors

discovered that chimeras with PCs had relatively larger brains, especially forebrains, and smaller bodies compared to a control group; in contrast, chimeras with ACs had the opposite phenotype, with smaller brains and larger bodies. These results suggested that genes expressed maternally and paternally contribute differently to the development of neural versus nonneural tissues in offspring. The experiments were designed such that the contribution of uniparental cells to particular tissues in the chimeras could be tracked from the expression of the bacterial reporter gene *lacZ*. Thus, the location of the ACs and PCs could be visualized in the brain for detailed observation. It was found that PCs localize preferentially in the neocortex, striatum, and hippocampus, but not the hypothalamus. In contrast, ACs contributed primarily to hypothalamic regions and preoptic areas, but not cortical areas (Barton et al., 1991; Fundele et al., 1990). These results indicated that imprinted genes expressed maternally and paternally may have differential functions in brain development, such that those expressed maternally may be required for the development of the neocortex, while those expressed paternally are associated with the development of hypothalamic areas. Subsequently, other studies have found that imprinted genes play diverse roles in brain development, including neuronal survival/apoptosis, neural differentiation, neurogenesis, neuronal migration, and axon outgrowth (Perez, Rubinstein, & Dulac, 2016).

In addition to neurodevelopment, imprinted genes have great significance in brain functions. Abnormal feeding behaviors are a common phenotype associated with imprinted gene mutations (Soellner et al., 2017). Indeed, imprinted genes regulate not only adult, but neonatal feeding as well. For example, deletion of *Peg3* reduces suckling and results in perinatal death (Curley et al., 2004), and mouse pups with paternal deletion

of G-protein α-subunit (*Gnas*) largely die before weaning because of impaired suckling (Plagge et al., 2004). Interestingly, *Gnas* is expressed in the motor areas that coordinate the muscles of the jaw and tongue in suckling behavior (Krechowec et al., 2012; Plagge et al., 2004). A study in humans found similar results, in that children with a paternal mutation in *GNAS* exhibited retarded growth and severe feeding problems (Bastepe, 2012). *Magel2*, a gene expressed paternally, and associated with PWS, plays an important role in infant feeding in the mouse. *Magel2*-deficient infants fail to attach to the nipple and suckle, which causes early death. When poor suckling was found to be attributable to the loss of *Magel2*, a single injection of oxytocin could correct the phenotype of mutant pups (Schaller et al., 2010). In addition, the maternally-biased *Asb4* gene is down-regulated during fasting periods in pro-opiomelanocortin (POMC) neurons, a group that is located in the arcuate nucleus of the hypothalamus, and regulates appetite and feeding control, and increased expression with administration of leptin, but not insulin (J. Y. Li, Chai, Zhang, Wang, & Mulholland, 2010; J. Y. Li et al., 2005). Conditional overexpression of *Asb4* in POMC neurons leads to increased food intake and energy expenditure, as well as reduced body weight (J. Y. Li et al., 2010).

1.3.5 Parental bias of allele expression and tissue-specific imprinting

Original studies defined genomic imprinting as complete silencing of one allele and monoallelic expression of the other. However, subsequent work has shown that absolute monoallelic expression is not the only form of genomic imprinting. The first imprinted gene found to have a parental allelic bias rather than monoallelic expression was *Ube3a*, an imprinted gene associated with AS. Albrecht et al. (Albrecht et al., 1997)

found that *Ube3a* was imprinted only in the brain and expressed the maternal allele over the paternal allele preferentially. Two months later, *Phlda2* was reported to demonstrate a maternal bias in expression in several human and mouse tissues (Qian et al., 1997).

In addition to maternal bias in allele expression, studies of *Ube3a* have demonstrated brain- and neuron-specific imprinting, which raises the question: Do parentally-biased imprinted genes express monoallelically in a tissue-specific manner or are they conserved across several tissues? Tissue-specific imprinting has been found in several imprinted genes. For example, Dechiara et al. (DeChiara et al., 1991) found that *Igf2* exhibits biallelic expression specifically in the leptomeninges and choroid plexus of the mouse brain. The mouse *Grb10* gene is expressed preferentially by the maternal allele in most tissues, while the paternal allele is expressed in the brain. *Grb10*'s tissue-specific imprinting is based on tissue-specific promoters (Arnaud et al., 2003; Hikichi, Kohda, Kaneko-Ishino, & Ishino, 2003). Imprinted genes, such as Delta Like Non-Canonical Notch Ligand 1 (*Dlk1*), exhibit selective loss of imprinting in the neurogenic niche astrocyte and neural stem cells in vivo (Ferron et al., 2011). In addition, next-generation sequencing, such as RNA sequencing, provides sensitive measures of expression to detect the genome-wide repertoire of parental allele bias. For example, one systematic characterization revealed that 28% of 82 imprinted genes for which expression of the parental alleles has been analyzed exhibited tissue-specific imprinting in at least two tissues (Prickett & Oakey, 2012). Brain is the adult tissue that exhibits the largest proportion of tissue-specific imprinting genes. More large-scale comprehensive screenings across different tissues are needed to answer the question above.

1.4 Random monoallelic expression (RME)

1.4.1 X chromosome in female

In addition to genomic imprinting, another major class of monoallelic expression is RME. RME leads to mosaicism of allelic expression in the organism, such that different cells express different alleles. The case of RME studied most is random X-chromosome inactivation (XCI), which silences one X chromosome in female cells randomly during development.

1.4.1.1 Random X-inactivation

Barr and Bertram (1949) reported the first evidence of XCI when they found that the sex of cat cells could be distinguished by a dense and small structure in the cell nuclei, referred to now as the "Barr body." A subsequent study by Ohno Kaplan, and Kinosita (Ohno, Kaplan, & Kinosita, 1959) indicated that the Barr body is a condensed X chromosome, and thereafter, Lyon (1961) proposed that the condensed X chromosome results from silencing one of the two X chromosomes, at random, during early development. We know now that two waves of XCI in mice occur during development (Augui, Nora, & Heard, 2011). The first takes place between the four to eight cell stage and results in imprinted paternal XCI (Adler, West, & Chapman, 1977; Kratzer & Gartler, 1978; Okamoto, Otte, Allis, Reinberg, & Heard, 2004). The silenced paternal X chromosome is reactivated at the blastocyst stage and then random XCI occurs (Okamoto et al., 2004). While the choice is made, the inactive state is preserved throughout cell mitosis.

During the past 20 years of XCI research, the region that regulates XCI, referred

to as X-inactivation center (Xic) (Rastan & Brown, 1990), has been identified and controls its major steps (Figure 1.3). The key component of the Xic is an lncRNA referred to as X-inactivation specific transcript (*Xist*), which produces a 17–20 kb RNA that decorates the X chromosome to trigger cis-inactivation during XCI (Brown et al., 1992). *Xist* has been found to be expressed only by the inactivated chromosome (Brown et al., 1991), and a deletion of *Xist* demonstrated that it is required for random and imprinted XCI (Marahrens, Panning, Dausman, Strauss, & Jaenisch, 1997; Penny, Kay, Sheardown, Rastan, & Brockdorff, 1996). Some factors that coordinate with Xist for XCI also have been identified. *Tsix*, an antisense repressor of *Xist*, was identified (J. T. Lee, Davidow, & Warshawsky, 1999), and is a lncRNA that represses *Xist* and coordinates X chromosome pairing to generate the epigenetic asymmetry required for random X inactivation (J. T. Lee & Lu, 1999). On the other hand, two lncRNAs, repeat A RNA (*RepA*) (Zhao, Sun, Erwin, Song, & Lee, 2008) and *Jpx* (Tian, Sun, & Lee, 2010), seem to function as *Xist* activators to promote XCI. The gene that encodes the E3 ubiquitin ligase RING finger 12 (RNF12), also known as RLIM, is located ~500 kb upstream of *Xist*, and its overexpression induces *Xist* RNA coating of the single X chromosome to cause partial derepression of *Xist* (Jonkers et al., 2009). Finally, the transcription factor Yin Yang 1 (YY1) has been shown to interact directly with *Xist* RNA and the *Xist* DNA locus, and serves as a bridge between the regulatory lncRNA and chromatin (Jeon & Lee, 2011).

For female-specific random XCI, two questions need to be addressed. First, what is the mechanism underlying the XCI in cells with more than one X chromosome? Several models of female-specific random XCI have been proposed, but the detailed

mechanism remains unclear (Goodrich, Panning, & Leung, 2016). Brown and colleagues' study found that *Xist* is activated only in female cells, and is expressed only by one X chromosome (Brown et al., 1991). How is one *Xist* allele marked for repression and the second allele in XX cells marked for activation? One possibility is that silencing the X results in a decrease in the amounts or activity of a dose-sensitive *Xist* activator. Several deletions on Xci were tested to identify the activator. Interestingly, deletion of 58 kb in Xic, including *Xist* and *Tsix*, is still able to trigger *Xist* to begin XCI during differentiation, suggesting that the key sequence associated with female-specific *Xist* activation should lie some distance from *Xist* (Monkhorst, Jonkers, Rentmeester, Grosveld, & Gribnau, 2008). A subsequent study proposed that Rnf12, located 500 kb upstream of *Xist*, is a dose-dependent activator of *Xist* expression in random XCI. Additional copies of mouse or human *RNF12* results in *Xist* up-regulation on the part of male and female embryonic stem cells (Jonkers et al., 2009). In contrast, random XCI decreased markedly in $Rnf12^{+/-}$ and $Rnf12^{-/-}$ female ESCs (Barakat et al., 2011). These results suggest that *Rnf12* may induce XCI to inactive one chromosome in a dose-dependent manner.

The second question related to female-specific random XCI is how is just one of the two X chromosomes chosen to be inactivated in females? The imbalance of *Tsix* and other activators' expression from two alleles may initiate random XCI. On the future active X chromosome, *Tsix* maintains expression from the X-inactivation intergenic transcription element to prevent loading of the *RepA*–PRC2 complex and the initiation of X-chromosome inactivation (Zhao et al., 2008). On the future inactive X chromosome, monoallelic loss of *Tsix* expression leads to the induction of the *Jpx* activator, which

allows *RepA*–PRC2 to promote *Xist* expression (Tian et al., 2010). Thereafter, *Xist* recruits PRC2 to a YY1-based nucleation center on the inactive chromosome, but is blocked from binding the active X (Jeon & Lee, 2011). Finally, an *Xist*–PRC2-ATRX complex spreads in cis across the X to silence the chromosome (Sarma et al., 2014). In addition to the factors and mechanisms mentioned above, several new players have been found to be involved in XCI. For example, SPEN, a large protein with several RNA-binding domains, was proposed to facilitate the initiation of XCI through direct binding to the Xist RNA (McHugh et al., 2015; Monfort et al., 2015). N6-methyladenosine modification on RNA, which is associated with RNA stability, translation, or splicing, was found to promote *Xist*-mediated transcriptional repression (Patil et al., 2016).

1.4.1.2 Nonrandom X-inactivation

During the first wave of XCI, the paternal X chromosome is imprinted and therefore silent during the four to eight cell stage. Further, random XCI patterns can be skewed in mature female somatic tissues, such that more cells choose one allele of the X chromosome than the other. Wu and colleagues' recent study revealed that X-inactivation is highly variable between individuals. The authors generated a dual-color, X-linked reporter system in combination with cell type-specific Cre drivers to visualize the XCI in different tissues and cell types (H. Wu et al., 2014). Surprisingly, females from the same litter exhibited strong variation, in that some littermates expressed the paternal allele (Xp) preferentially, while others largely expressed the maternal allele (Xm). In addition to individual variation, regional and cell-type variation in XCI mosaicism has been found in different tissues. Thus, it appears that XCI fluctuates greatly, rather than showing an even

distribution of equal numbers of Xm- and Xp-expressing cells.

The parent-of-origin effect can skew X inactivation, and animal models have revealed the existence of locus-specific imprinting of X-linked genes in the developing brain (Raefski & O'Neill, 2005). Several X-linked genes also were characterized as imprinted genes (Davies et al., 2005; Kobayashi et al., 2006; Raefski & O'Neill, 2005). Some X-linked imprinted genes showed tissue-specific and developmental-specific imprinting. For example, *Xlr3b*, the X-linked imprinted gene discovered first, exhibits strong repression of the paternal allele in neonatal liver, while it exhibits biallelic expression in placenta. *Xlr3b* expresses both alleles in diencephalon and hindbrain at E14.5, but paternal repression appears at birth and is maintained into adulthood (Raefski & O'Neill, 2005). In addition to individual genes, recent evidence using next generation sequencing has shown that preferential paternal XCI occurs in the neonatal brain (Gregg, Zhang, Butler, Haig, & Dulac, 2010; X. Wang, Soloway, & Clark, 2010). However, mechanisms underlying different parent-of-origin effects on XCI are still unclear.

1.4.1.3 X-inactivation escapees

Some genes can escape X inactivation and are expressed from both alleles. A systematic survey has shown that approximately 15% of X-linked genes escape in humans (Carrel & Willard, 2005), and a recent study even showed that at least 23% of X-linked genes are influenced by incomplete XCI (Tukiainen et al., 2016). In contrast, only a few X-linked genes (3%) escape in the mouse, as demonstrated in an RNA-seq study in a somatic cell line derived from a hybrid mouse (Yang, Babak, Shendure, & Disteche, 2010). X escapees are found throughout the X chromosome, but are located

predominantly in pseudoautosomal regions (PAR). PAR is a small region with homologous sequences of nucleotides on the X and Y chromosomes, and genes in the PARs express equivalent alleles on X and Y chromosomes in males, and on both X chromosomes in females. Interestingly, mouse PAR has been found to be much shorter than human PAR (Perry, Palmer, Gabriel, & Ashworth, 2001). Thus, the location of escapees along the X chromosome differs between humans and mice. Escapees in humans are clustered, while in mice, single escape genes lie in the regions of silenced chromatin (Carrel & Willard, 2005; Yang et al., 2010). This major difference in X-inactivation escape between mice and humans suggests that different mechanisms are involved in X-inactivation in the two species. Importantly, XCI escape varies among individuals. In humans, 15% of genes were found to be escaped in all samples, while 20% were inactivated in some, but not all samples (Carrel & Willard, 2005). Further, escape from X-inactivation can change developmentally. For example, Kdm5c escapes in adult mouse tissues, including liver, lung, and brain, but not in embryonic tissues (Lingenfelter et al., 1998).

The mechanisms involved in escaping XCI still have not been characterized fully. The distribution of repeat elements may be important in escape and XCI. Higher concentrations of long, interspersed nuclear elements-1 (LINE-1) repeats occur near genes that are subject to inactivation, while escapees have fewer LINE-1 repeats (Bailey, Carrel, Chakravarti, & Eichler, 2000). The following study revealed LINEs' function in creating a silent nuclear compartment during XCI (Chow et al., 2010). Other sequences and motifs also are associated with the escape from XCI, for example, long terminal repeats (LTRs; Tsuchiya et al., 2004) or AT-rich motifs (Z. Wang, Willard, Mukherjee,

& Furey, 2006). In addition to DNA sequence, CTCF, a key insulator protein, plays a role in XCI's escape. CTCF binds uniquely to transcription start sites for X-linked escapees. It also has been found that the CpG island at the 5' end of *Kdm5c* remains hypomethylated throughout mouse development, which may be attributable to CTCF binding (Filippova et al., 2005). However, some evidence has shown that CTCF-binding alone might not be sufficient to protect silencing. For example, insertion of CTCF-binding sites is not sufficient to cause escape from XCI (Ciavatta, Kalantry, Magnuson, & Smithies, 2006). Taken together, it is clear that further studies are required to identify a common mechanism of escape from X inactivation.

1.4.2 Known autosomal RME

In addition to random X inactivation in females, RME has been detected for some autosomal genes. This phenomenon has been studied since the 1960s and has been found to affect specific gene families in the central nervous and immune systems, for example, antigen receptors/immunoglobulins in B and T cells, the olfactory receptor gene family, and protocadherins in nervous system. The RME of these three gene families is reviewed below.

1.4.2.1 Immunoglobulin (Ig)

Autosomal RME was first reported for antigen receptors (Pernis, Chiappino, Kelus, & Gell, 1965). One of the well-known features of B and T lymphocytes, regarded as Burnet's clonal selection theory, is that each B and T cell recognizes only one antigen and explains the pathogen-specific production of antibodies (Burnet, 1959). This "one

cell–one antibody" rule signifies that antigen receptors in individual B cells contain only one particular antigen-binding site that corresponds to the respective pathogen-associated antigens and results in a highly-specific antibody response. Indeed, expression of additional functional antigen receptors could induce multiple specificities of immune cells and lead to deleterious phenotypes such as those with autoimmune disorders (Pelanda, 2014). The genetic mechanism underlying "one cell–one antibody" is allelic exclusion, a process of monoallelic rearrangement of Ig gene segments (Brady, Steinel, & Bassing, 2010). The rearrangement also is referred to as VDJ recombination, a process that assembles the V, D, and J gene segments that encode the variable region of the antigen receptor molecule randomly to generate unique antigen receptors. During B and T cell development, one of the two alleles undergoes various VDJ combinations, but only a fraction of the resulting Ig genes is functional and able to assemble successfully as a surface-expressed antigen receptor. If rearrangement assembles a functional receptor, further rearrangements are blocked by a feedback inhibition mechanism. However, if rearrangement fails to produce a functional receptor, rearrangement continues at the original allele or the other allele. Allelic exclusion is thought to be a special case. Most RME cases, such as X-linked genes in females, the olfactory receptor family, and most imprinting genes are regulated epigenetically, such that only one of the two alleles is transcribed. Conversely, antigen receptor loci are expressed from both alleles, but only the functional allele will translate into protein. This conclusion was drawn from a study in which mice were engineered to carry two different, prearranged functional *IgH* and *Igk* alleles, both of which were translated and expressed on the surface of mature B cells (Fraenkel et al., 2007; Sonoda et al., 1997). Thus, the mechanisms of allelic exclusion

involve a unique rearrangement process.

Several working models of monoallelic rearrangement and mechanisms have been identified that ensure the recombination machinery acts on the allele chosen (Vettermann & Schlissel, 2010). One mechanism that labels the expressed versus silenced allele is asynchronous replication, in which replication of the expressed allele occurs earlier than that of the silent allele during the S phase. Genes located within asynchronously replicating regions are monoallelic, such as imprinted genes (Simon et al., 1999), the olfactory receptor family (Chess, Simon, Cedar, & Axel, 1994), and the inactive X chromosome (Priest, Heady, & Priest, 1967). Antigen receptor loci exhibit asynchronous replication as well (Mostoslavsky et al., 2001). In mature B cells, the rearranged *Igh* and *Igk* alleles replicate early in S phase, while the un-rearranged alleles replicate later. In pro-B- and pre-B-cell stages, the early replicating allele of *Igk* can predict the one chosen for rearrangement when the cells initiate differentiation (Farago et al., 2012). In addition, different chromatin modifications likely cooperate to modulate the chromatin accessibility to recombination-activating genes 1 and 2 (*RAGs*), the enzymes essential for VDJ rearrangement and allelic exclusion. The future rearranged *Igk* allele is marked with gene-activating histone marks, such as acetylation on H3 histone or H3K4me3. In contrast, the opposing allele is localized to heterochromatin, which maintains genes in a repressive state (Farago et al., 2012; Goldmit et al., 2005). The presence of H3K4me3 can act as a docking site to recruit the *RAG2* enzyme for future rearrangement (Ji et al., 2010). DNA methylation plays a role in VDJ recombination as well, and has been shown to reduce the activity of the *RAG* proteins significantly; further, reduction of DNA methylation levels can promote rearrangement in vitro (Ji, Zhang, Lee, Cedar, &

Bergman, 2003; Nakase, Takahama, & Akamatsu, 2003). Thus, these results demonstrate that allelic exclusion on VDJ loci involves a complex epigenetic mechanism that results in monospecificity of B and T cells.

1.4.2.2 Clustered protocadherins

Clustered protocadherins are another gene family that demonstrates RME (Chess, 2005; Esumi et al., 2005; Kaneko et al., 2006). Protocadherins (Pcdhs) are the largest mammalian subgroup of the cadherin superfamily and are expressed predominantly in the nervous system (Q. Wu & Maniatis, 1999). Based on genomic organization, Pcdhs have been defined according to two subgroups: nonclustered and clustered Pcdhs. Nonclustered Pcdhs, known as δ-Pcdhs, are scattered throughout the genome. The clustered Pcdhs contain three gene clusters, Pcdha, Pcdhb, and Pcdhg, including approximately 60 genes in humans and mice (Q. Wu & Maniatis, 1999; Q. Wu et al., 2001), and the genomic organization of clustered Pcdhs is conserved highly in both species (Yagi, 2008). Pcdha and Pcdhg are composed of multiple variable exons that encode the extracellular, transmembrane, and cytoplasmic domains of the protein, as well as three constant exons that encode a shared intracellular region. Pcdhb is composed only of variable exons (Tasic et al., 2002; X. Wang et al., 2002). Diverse clustered Pcdhs isoforms are expressed combinatorially in individual neurons to provide them a unique identity for neuronal connectivity and interneuronal recognition (Yagi, 2012). Complete deletion of clustered Pcdhs in cortical neurons was shown to produce impaired formation and stabilization of neuronal connections (Hasegawa et al., 2016). In particular, they appeared to allow neurons to recognize “self,” and thus prevented them from forming

synapses on their own processes (Lefebvre, Kostadinov, Chen, Maniatis, & Sanes, 2012).

Unlike rearrangement in antigen receptors, the mechanism of stochastic expression of clustered Pcdhs has been coined “promoter choice” (Tasic et al., 2002). Each variable exon of Pcdhs contains its own promoter, which includes a conserved sequence element, and the transcription is initiated from a unique promoter. When the promoter of the variable exon is chosen stochastically, a long transcript through the rest of the cluster is spliced subsequently, probably through interaction of the downstream HS7 and HS5-1 sequence, to produce a mature mRNA that contains only the 5’-most variable exons and the three constant exons (Ribich, Tasic, & Maniatis, 2006; Yokota et al., 2011). Both PCDHα and PCDHγ, encoded by *Pcdha* and *Pcdhg*, follow this transcriptional and splicing process. However, the detailed mechanisms of the way in which one promoter is chosen stochastically over another remain unclear. The insulator CTCF and nuclear phosphoprotein RAD21 may play roles in regulating random promoter choice. Active *Pcdha* promoters are bound to both CTCF and the nuclear RAD21, as is the HS5-1 enhancer (Kehayova, Monahan, Chen, & Maniatis, 2011; Monahan et al., 2012). Pcdhs may be inactivated by DNA methylation as well. Pcdhs that express stochastically and monoallelically in neurons exhibit mixed DNA methylation patterns in their promoter, while promoters in which Pcdhs are expressed constitutively and biallelically are hypomethylated (Kawaguchi et al., 2008). In addition to stochastic promoter choice, the mechanisms that stabilize the monoallelic expression of Pcdhs are unclear.

1.4.2.3 Ordorant receptor (OR)

Similar to antigen receptors, which exhibit the "one antibody–one cell" principle to guarantee appropriate immune function, the OR also follows a "one OR–one neuron" rule to ensure proper function and wiring of olfactory systems (Serizawa, Miyamichi, & Sakano, 2004). There are approximately 350 human OR genes and 1400 mouse OR genes (Godfrey, Malnic, & Buck, 2004; X. Zhang & Firestein, 2002; Zozulya, Echeverri, & Nguyen, 2001), which are expressed primarily in the olfactory sensory neurons (OSNs) in the olfactory epithelium. Only one OR allele among hundreds is expressed in each OSN (Chess et al., 1994; Ishii et al., 2001). Originally, monoallelic expression of ORs was proposed based on transcript cloning experiments (Chess et al., 1994). Recently, transcriptome analysis of single cells from hybrid mice demonstrated that only one allele of an OR was expressed in most cells analyzed (Saraiva et al., 2015).

The regulation of monogenic and monoallelic expression of ORs involves many epigenetic mechanisms. Asynchronous replication, a DNA replication process mentioned above, was observed in ORs. (Singh et al., 2003). The mechanism that allows each mouse olfactory neuron to express only one of 1400 ORs stochastically has been elucidated. Lomvadas et al. found that a single enhancer of an OR gene cluster interacts with multiple OR gene promoters, suggesting that, to be activated, ORs located on different chromosomes compete for a single enhancer (Lomvardas et al., 2006). Differential marking of the two alleles with repressive and active markers for transcription, respectively, was found to function in monoallelic expression of the OR genes. Remarkably, genome-wide chromatin analysis of the olfactory epithelium found histone modifications of OR genes consisting of H3K9me3 and H4K20me3, which are associated

with constitutive heterochromatin. H3K9me3 and H4K20me3 cover the OR gene coding regions and intergenic regions to form a constitutive heterochromatin block that prevents transcription factors from binding (Magklara et al., 2011). While the OR is chosen and activated, the expressed OR allele lacks these repressive markers, and is marked instead by H3K4me3, an active marker for transcription (Magklara et al., 2011). These results suggest that OR activation involves an epigenetic switch from a repressive to an active epigenetic signature. LSD1, a histone demethylase, was found to be essential for OR gene activation (Lyons et al., 2013). Deletion of LSD1 in OSN prior to OR choice results in a complete loss of OR expression, while it has no effect after OR choice.

After a single OR allele is activated, further mechanisms prevent the activation of additional OR alleles and maintain monoallelic expression. This mechanism involves negative feedback during translation of the OR at high levels in the endoplasmic reticulum **(**ER) to trigger the unfolded protein response, a cellular stress response. The unfolded protein response induces expression of Adenylyl cyclase type 3 (*Adcy3*) to downregulate LSD1 and prevent demethylation (Dalton, Lyons, & Lomvardas, 2013). Another mechanism that maintains the monoallelic expression of a chosen OR in an individual neuron also was described recently (Abdus-Saboor et al., 2016). The study forced ectopic expression of another OR gene in mature OSNs and found that it suppressed the expression of the endogenous OR gene chosen previously. This postselection refinement demonstrated that the endogenous OR that is transcribed at a higher level of expression is suppressed less efficiently compared to an endogenous OR with lower expression. This result revealed a competitive relationship between OR alleles and a negative feedback system that ensures the singularity of OR allelic expression.

1.4.3 RME studied by genome-wide approaches

1.4.3.1 Studies from cell lines

In addition to the gene families mentioned above that show RME, a small number of isolated autosomal genes with RME also have been identified, for example, interleukin genes in T cells (Hollander et al., 1998; Kelly & Locksley, 2000), *Tlr4* in B cells (Pereira, Girard, Chaby, Cumano, & Vieira, 2003), glial fibrillary acidic protein (*Gfap*) gene in astrocytes (Takizawa, Gudla, Guo, Lockett, & Misteli, 2008), and forkhead box protein P2 (*FOXP2*) gene in human lymphoblasts (Adegbola et al., 2015). These findings indicated that other genes with RME might exist in the genome. With the advance of whole genome sequencing, which provides resolution at the level of single nucleotide polymorphisms and the use of new methods, such as single nucleotide polymorphism (SNP) genotyping arrays, and next-generation sequencing, genome-wide studies of RME genes have been conducted. The earliest genome-wide analysis of RME was performed in Chess' laboratory, in which SNP arrays were performed to examine cDNA reverse-transcribed from bulk RNA of human monoclonal B-lymphoblastoid cell lines (Gimelbrant, Hutchinson, Thompson, & Chess, 2007). This seminal study estimated that nearly 5-10% of approximately 4000 autosomal genes analyzed were subject to RME. A later SNP microarray study from the same group found a similar degree of RME in immortalized mouse lymphoblasts, but fewer (~2%) in mouse fibroblast (Zwemer et al., 2012). More recently, RNA-seq also was performed to determine the extent of RME in immortalized clones of mouse tail-tip fibroblasts and embryonic fibroblast (Pinter et al., 2015). In addition, immortalized clones of human and mouse cells have been used to explore epigenetic features related to RME (Nag et al., 2013; Nag, Vigneau, Savova,

Zwemer, & Gimelbrant, 2015).

Cell culture systems have been used widely in genome-wide studies of RME genes, principally because of a technical challenge in which bulk RNA-seq analysis of polyclonal cell population pooled by the cells that express either the maternal or paternal allele of an RME gene, such as in tissues, would make expression appear biallelic for every RME gene (Nag et al., 2013). Therefore, monoclonal samples derived from a single cell in a cell culture system were used for the analysis. However, analyses of cell culture systems, and especially of immortalized cell lines, can be confounded by copy number variation (Lucito et al., 2003) or chromosomal abnormalities (Gaztelumendi & Nogues, 2014; Maitra et al., 2005). These confounding factors should be scrutinized carefully when analyzing results from cell culture systems.

1.4.3.2 Studies from stem cells

In addition to immortalized cell lines, RME in stem cells has been examined with genome-wide approaches. A stem cell is an undifferentiated cell type capable of giving rise to other kinds of cell following differentiation, and therefore, it can be used to observe RME in different cell lineages. Originally, an SNP-sensitive microarray identified 2-2.4% of analyzed genes subject to RME in mouse and human neural stem cells (Jeffries et al., 2012; S. M. Li et al., 2012). Two recent studies co-published by two different groups have extended our understanding of RME greatly and explored the change of RME upon cell differentiation. These studies used bulk RNA-seq analysis of clonally expanded stem cell populations derived from hybrid mice, and identified ~0.5-1% of analyzed autosomal genes in embryonic stem cells, and 5–10% of genes in neural

progenitor cells (Eckersley-Maslin & Spector, 2014; Gendrel et al., 2014). These results suggest that RME is acquired upon lineage commitment.

1.4.3.3 Studies from single cell sequencing

Single cell sequencing examines individual cells with optimized next generation sequencing technologies that provide a higher resolution of cellular variation. Disassociating cells from tissues to analyze individual cells is another approach that bypasses the technical challenge in analysis of polyclonal cell pooling. The first study that employed single cell RNA-seq on blastomeres from preimplantation embryos in crossed mouse strains identified 12–24% of genes on autosomes as RME across all preimplantation stages analyzed (Deng, Ramskold, Reinius, & Sandberg, 2014). A similar proportion of RME was observed in single-cell RNA-seq analysis of mouse hepatocytes and fibroblasts. The authors found that allelic expression patterns rarely were maintained across cells that divided in the same embryo; therefore, the authors proposed dynamic RME (Reinius & Sandberg, 2015), which differs conceptually from clonally-inherited RME, and is thought to derive from transient stochastic monoallelic expression. In addition to mouse blastomeres, mouse neural progenitor cells (Marinov et al., 2014), human primary fibroblasts (Borel et al., 2015), and human T cells (Reinius et al., 2016) have been examined by single cell RNA-seq in different studies. Although single cell transcriptome profiling has better resolution that allows observation of the dynamics of transcriptional states, technical noise, such as random allele dropout and variable capture efficiency, may inflate allelic calling and confound estimates of RME genes (Benitez, Cheng, & Deng, 2017). The technical control of experiments should be addressed

carefully to estimate RME accurately (J. K. Kim, Kolodziejczyk, Ilicic, Teichmann, & Marioni, 2015).

1.4.3.4 Potential mechanisms underlying isolated RME

Several studies have explored various epigenetic features associated with isolated RME. Chromatin immunoprecipitation followed by genome-wide sequencing on murine neural progenitor cells was found to transcribe an allele enriched for an active (H3K4me2 and/or H3K4me3) marker, while it silenced an allele enriched for repressive (H3K9me3) markers (Eckersley-Maslin & Spector, 2014). Nag and colleagues identified the chromatin signature, H3K36me3, an active histone mark for transcription, and H3K27me3, a repressive histone mark for silencing, across the bodies of many RME genes in human (Nag et al., 2013) and mouse cells (Nag et al., 2015). The same group used this dual active/inactive chromatin signature subsequently to classify RME genes in different tissues for further analysis (Savova, Chun, et al., 2016; Savova, Patsenker, Vigneau, & Gimelbrant, 2016; Savova, Vinogradova, Pruss, Gimelbrant, & Weiss, 2017), but the accuracy of this predictive tool still needs to be validated. Bisulfate sequencing used to analyze DNA methylation on RME genes exhibited varied results, and showed that the allele expressed in some genes had decreased levels of DNA methylation, while others did not (Gendrel et al., 2014). Administration of 5-azacytidine, a DNA methyltransferase inhibitor, to lower DNA methylation could not increase biallelic expression of monoallelic genes (Eckersley-Maslin & Spector, 2014; Gendrel et al., 2014), suggesting that DNA methylation is not a widespread marker that regulates RME. Recently, an allele-specific assay for transposase-accessible chromatin with high

throughput sequencing (ATAC-Seq), a technique used to study chromatin accessibility, was performed on mouse embryonic stem cells and neural progenitor cells to profile active regulatory DNA sequences across the genome to identify regulatory elements of RME (Xu et al., 2017). Interestingly, regulatory elements of RME were enriched at promoters, while elements of genetically-determined monoallelic expression tended to be at enhancers, suggesting that regulation of RME is associated with a gene's promoter rather than with long-range regulatory elements. Taken together, these results suggest that no single epigenetic marker can explain RME, and indicate that multiple epigenetic regulations likely work together to maintain RME.

1.5 Unanswered questions and dissertation goals

Genetic risk factors for complex diseases, such as autism spectrum disorders (ASDs), have been challenging to define, and are an obstacle in understanding the basis of these disorders and devising effective therapeutic strategies. Several recent genomic studies revealed that loss-of-function heterozygous mutations, rather than recessive mutations, are involved in ASDs, as well as schizophrenia. However, we do not understand fully the way in which heterozygous mutations contribute to disease. In a diploid genome, most genes have two copies, and the healthy backup copy can reduce the effect of heterozygous mutations. However, if the mutated allele is the only allele expressed because of epigenetic effects that silence the expression of the healthy allele, then a heterozygous mutation will have a stronger influence. Epigenetic effects involve modifications to the genome that alter gene expression without changing the genome sequence. However, we know little about the prevalence of ASEs in vivo for most genes

in the genome and how their effect influences complex disorders.

ASEs are known from genomic imprinting in which either the maternal or paternal allele is silenced for a small number of genes in the genome because of epigenetic effects that are inherited from the mother or father. These effects influence the genetics of heterozygous mutations, for example, the same mutation on chromosome 15 can cause Prader-Willi Syndrome or Angelman Syndrome depending on the parental origin of the mutation. RME effects also are known to exist in the genome and affect genes on the X-chromosome in females, immune genes, protocadherins, and olfactory receptors. Recently, some autosomal genes with RME were found scattered in the genome in vitro.

Currently, we know little about RME genes in vivo, as robust and scalable methods to profile RME genes in tissues have not been developed. Further, the spatial and temporal variability of ASE genes, such as imprinted and RME genes, have not been characterized fully. It also is unclear whether the mutation of RME genes influences genetic architecture in vivo. Therefore, the objective of the dissertation is focused on: (1) developing methods to profile allele-specific expression effects in vivo, and (2) analyzing cellular expression of RME genes in heterozygous mutated mice.

- **Goal 1: Develop methods to profile ASE genes in vivo**. Previous genome-wide profiles of RME that used materials collected from cell culture systems has expanded our understanding of RME. However, chromosome abnormality and gene copy variation that occur typically in cultured cells may confound estimates of RME. Thus, the first goal of the dissertation was to develop accurate approaches to identify RME genes in vivo. These results are described in Chapter

3.

- **Goal 2: Genome-wide analysis of ASE genes in different tissues and developmental stages**. Some imprinted genes have been shown to express in a tissue-specific manner, for example, imprinting of Ube3a occurs only in the brain. Whether parental-biased genes with differential expression between two alleles also are exhibited in a similar manner in different tissues has not been investigated comprehensively, and the temporal and spatial process of RME in vivo remains unclear as well. Thus, the dissertation's second goal was to identify ASE genes in different tissues and different developmental stages in vivo and validate the results using independent approaches. The results of these experiments are described in Chapters 2 and 3.
- **Goal 3: Determine whether RME genes affect the cellular expression of mutated and healthy alleles for inherited heterozygous mutations**. The parental origin of the deleted imprinted region of chromosome 15 results in the different phenotypes known as Prader-Willi syndrome and Angelman syndrome; however, the genetic architecture underlying the mutation of RME genes in vivo has not yet been explored. Therefore, to address the third goal, cell subpopulations that express the mutant versus the wildtype allele of the RME gene were examined in vivo using heterozygous reporter mice. The results of these experiments are described in Chapter 2.

1.6 References

Abdus-Saboor, I., Al Nufal, M. J., Agha, M. V., de Brimont, M. R., Fleischmann, A., & Shykind, B. M. (2016). An expression refinement process ensures singular odorant receptor gene choice. *Current Biology*, *26*(8), 1083-1090.

Adegbola, A. A., Cox, G. F., Bradshaw, E. M., Hafler, D. A., Gimelbrant, A., & Chess, A. (2015). Monoallelic expression of the human FOXP2 speech gene. *Proceedings of the National Academy of Sciences*, *112*(22), 6848-6854.

Adler, D. A., West, J. D., & Chapman, V. M. (1977). Expression of alpha-galactosidase in preimplantation mouse embryos. *Nature, 267*(5614), 838-839.

Albrecht, U., Sutcliffe, J. S., Cattanach, B. M., Beechey, C. V., Armstrong, D., Eichele, G., & Beaudet, A. L. (1997). Imprinted expression of the murine Angelman syndrome gene, Ube3a, in hippocampal and Purkinje neurons. *Nature Genetics*, *17*(1), 75-78.

Allen, N. D., Logan, K., Lally, G., Drage, D. J., Norris, M. L., & Keverne, E. B. (1995). Distribution of parthenogenetic cells in the mouse brain and their influence on brain development and behavior. *Proceedings of the National Academy of Sciences, 92*(23), 10782-10786.

Allfrey, V. G., Faulkner, R., & Mirsky, A. E. (1964). Acetylation and methylation of histones and their possible role in the regulation of RNA syndissertation. *Proceedings of the National Academy of Sciences, 51*, 786-794.

Arnaud, P., Monk, D., Hitchins, M., Gordon, E., Dean, W., Beechey, C. V., ... Moore, G. E. (2003). Conserved methylation imprints in the human and mouse GRB10 genes with divergent allelic expression suggests differential reading of the same mark. *Human Molecular Genetics*, *12*(9), 1005-1019.

Augui, S., Nora, E. P., & Heard, E. (2011). Regulation of X-chromosome inactivation by the X-inactivation centre. *Nature Reviews. Genetics*, *12*(6), 429.

Bailey, J. A., Carrel, L., Chakravarti, A., & Eichler, E. E. (2000). Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: The Lyon repeat hypodissertation. *Proceedings of the National Academy of Sciences, 97*(12), 6634-6639.

Barakat, T. S., Gunhanlar, N., Pardo, C. G., Achame, E. M., Ghazvini, M., Boers, R., ... Gribnau, J. (2011). RNF12 activates Xist and is essential for X chromosome inactivation. *PLoS Genetics*, *7*(1), e1002001.

Barlow, D. P., & Bartolomei, M. S. (2014). Genomic imprinting in mammals. *Cold Spring Harbor Perspectives in Biology*, *6*(2), a018382.

Barlow, D. P., Stoger, R., Herrmann, B. G., Saito, K., & Schweifer, N. (1991). The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. *Nature, 349*(6304), 84-87.

Barr, M. L., & Bertram, E. G. (1949). A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein syndissertation. *Nature, 163*(4148), 676.

Bartolomei, M. S., Zemel, S., & Tilghman, S. M. (1991). Parental imprinting of the mouse H19 gene. *Nature, 351*(6322), 153-155.

Barton, S. C., Ferguson-Smith, A. C., Fundele, R., & Surani, M. A. (1991). Influence of paternally imprinted genes on development. *Development, 113*(2), 679-687.

Bastepe, M. (2012). Relative functions of Gαs and its extra-large variant XLαs in the endocrine system. *Hormone and Metabolic Research*, *44*(10), 732-740.

Benitez, J. A., Cheng, S., & Deng, Q. (2017). Revealing allele-specific gene expression by single-cell transcriptomics. *The International Journal of Biochemistry & Cell Biology, 90*, 155-160.

Bergman, Y., & Cedar, H. (2013). DNA methylation dynamics in health and disease. *Nature Structural & Molecular Biology*, *20*(3), 274-281.

Bird, A. P. (1986). CpG-rich islands and the function of DNA methylation. *Nature, 321*(6067), 209-213.

Borel, C., Ferreira, P. G., Santoni, F., Delaneau, O., Fort, A., Popadin, K. Y., ... Padioleau, I. (2015). Biased allelic expression in human primary fibroblast single cells. *The American Journal of Human Genetics*, *96*(1), 70-80.

Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B., & Bestor, T. H. (2001). Dnmt3L and the establishment of maternal genomic imprints. *Science, 294*(5551), 2536-2539.

Brady, B. L., Steinel, N. C., & Bassing, C. H. (2010). Antigen receptor allelic exclusion: An update and reappraisal. *The Journal of Immunology*, *185*(7), 3801-3808.

Brown, C. J., Ballabio, A., Rupert, J. L., Lafreniere, R. G., Grompe, M., Tonlorenzi, R., & Willard, H. F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature, 349*(6304), 38-44.

Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafreniere, R. G., Xing, Y., Lawrence, J., & Willard, H. F. (1992). The human XIST gene: Analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell, 71*(3), 527-542.

Buiting, K., Saitoh, S., Gross, S., Dittrich, B., Schwartz, S., Nicholls, R. D., & Horsthemke, B. (1995). Inherited microdeletions in the Angelman and Prader–Willi syndromes define an imprinting centre on human chromosome 15. *Nature Genetics*, *9*(4), 395-400.

Burnet, F. M. (1959). *The clonal selection theory of acquired immunity*. Nashville, TN: Vanderbilt University Press.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., . . . Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science, 298*(5595), 1039-1043.

Carrel, L., & Willard, H. F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature, 434*(7031), 400-404.

Cassidy, S. B., Dykens, E., & Williams, C. A. (2000). Prader-Willi and Angelman syndromes: Sister imprinted disorders. *American Journal of Medical Genetics Part A*, *97*(2), 136-146.

Cattanach, B. M., & Kirk, M. (1985). Differential activity of maternally and paternally derived chromosome regions in mice. *Nature, 315*(6019), 496-498.

Chen, K., Zhao, B. S., & He, C. (2016). Nucleic acid modifications in regulation of gene expression. *Cell Chemical Biology*, *23*(1), 74-85.

Chernukhin, I. V., Shamsuddin, S., Robinson, A. F., Carne, A. F., Paul, A., El-Kady, A. I., ... Klenova, E. M. (2000). Physical and functional interaction between two pluripotent proteins, the Y-box DNA/RNA-binding factor, YB-1, and the multivalent zinc finger factor, CTCF. *Journal of Biological Chemistry*, *275*(38), 29915-29921.

Chess, A. (2005). Monoallelic expression of protocadherin genes. *Nature Genetics*, *37*(2), 120-121.

Chess, A. (2013). Random and non-random monoallelic expression. *Neuropsychopharmacology*, *38*(1), 55.

Chess, A. (2016). Monoallelic gene expression in mammals. *Annual Review of Genetics*, *50*, 317-327.

Chess, A., Simon, I., Cedar, H., & Axel, R. (1994). Allelic inactivation regulates olfactory receptor gene expression. *Cell, 78*(5), 823-834.

Chow, J. C., Ciaudo, C., Fazzari, M. J., Mise, N., Servant, N., Glass, J. L., . . . Heard, E. (2010). LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. *Cell, 141*(6), 956-969.

Ciavatta, D., Kalantry, S., Magnuson, T., & Smithies, O. (2006). A DNA insulator prevents repression of a targeted X-linked transgene but not its random or imprinted X inactivation. *Proceedings of the National Academy of Sciences, 103*(26), 9958-9963.

Ciccone, D. N., Su, H., Hevi, S., Gay, F., Lei, H., Bajko, J., . . . Chen, T. (2009). KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. *Nature, 461*(7262), 415-418.

Cowley, M., Garfield, A. S., Madon-Simon, M., Charalambous, M., Clarkson, R. W., Smalley, M. J., ... Oakey, R. J. (2014). Developmental programming mediated by complementary roles of imprinted Grb10 in mother and pup. *PLoS Biology*, *12*(2), e1001799.

Curley, J. P., Barton, S., Surani, A., & Keverne, E. B. (2004). Coadaptation in mother and infant regulated by a paternally expressed imprinted gene. *Proceedings of the Royal Society B: Biological Sciences*, *271*(1545), 1303.

Dalton, R. P., Lyons, D. B., & Lomvardas, S. (2013). Co-opting the unfolded protein response to elicit olfactory receptor feedback. *Cell, 155*(2), 321-332.

Davies, W., Isles, A., Smith, R., Karunadasa, D., Burrmann, D., Humby, T., ... Wilkinson, L. (2005). Xlr3b is a new imprinted candidate for X-linked parent-of-origin effects on cognitive function in mice. *Nature Genetics*, *37*(6), 625.

Davis, T. L., Trasler, J. M., Moss, S. B., Yang, G. J., & Bartolomei, M. S. (1999). Acquisition of theH19Methylation imprint occurs differentially on the parental alleles during spermatogenesis. *Genomics*, *58*(1), 18-28.

Day, J. J., & Sweatt, J. D. (2010). DNA methylation and memory formation. *Nature Neuroscience*, *13*(11), 1319-1323.

Day, T., & Bonduriansky, R. (2004). Intralocus sexual conflict can drive the evolution of genomic imprinting. *Genetics, 167*(4), 1537-1546.

DeChiara, T. M., Robertson, E. J., & Efstratiadis, A. (1991). Parental imprinting of the mouse insulin-like growth factor II gene. *Cell, 64*(4), 849-859.

Deng, Q., Ramskold, D., Reinius, B., & Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science, 343*(6167), 193-196.

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., . . . Gingeras, T. R. (2012). Landscape of transcription in human cells. *Nature, 489*(7414), 101-108.

Eckersley-Maslin, M. A., & Spector, D. L. (2014). Random monoallelic expression: regulating gene expression one allele at a time. *Trends in Genetics*, *30*(6), 237-244.

Esumi, S., Kakazu, N., Taguchi, Y., Hirayama, T., Sasaki, A., Hirabayashi, T., ... Yagi, T. (2005). Monoallelic yet combinatorial expression of variable exons of the protocadherin-[alpha] gene cluster in single neurons. *Nature Genetics*, *37*(2), 171.

Farago, M., Rosenbluh, C., Tevlin, M., Fraenkel, S., Schlesinger, S., Masika, H., . . . Bergman, Y. (2012). Clonal allelic predetermination of immunoglobulin-kappa rearrangement. *Nature, 490*(7421), 561-565.

Ferguson-Smith, A. C. (2011). Genomic imprinting: The emergence of an epigenetic paradigm. *Nature Reviews. Genetics*, *12*(8), 565.

Ferguson-Smith, A. C., Cattanach, B. M., Barton, S. C., Beechey, C. V., & Surani, M. A. (1991). Embryological and molecular investigations of parental imprinting on mouse chromosome 7. *Nature, 351*(6328), 667-670.

Ferron, S. R., Charalambous, M., Radford, E., McEwen, K., Wildner, H., Hind, E., . . . Ferguson-Smith, A. C. (2011). Postnatal loss of Dlk1 imprinting in stem cells and niche astrocytes regulates neurogenesis. *Nature, 475*(7356), 381-385.

Filippova, G. N., Cheng, M. K., Moore, J. M., Truong, J. P., Hu, Y. J., Tsuchiya, K. D., & Disteche, C. M. (2005). Boundaries between chromosomal domains of X inactivation and escape bind CTCF and lack CpG methylation during early development. *Developmental Cell*, *8*(1), 31-42.

Filippova, G. N., Fagerlie, S., Klenova, E. M., Myers, C., Dehner, Y., Goodwin, G., ... Lobanenkov, V. V. (1996). An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Molecular and Cellular Biology*, *16*(6), 2802-2813.

Fournier, C., Goto, Y., Ballestar, E., Delaval, K., Hever, A. M., Esteller, M., & Feil, R. (2002). Allele-specific histone lysine methylation marks regulatory regions at imprinted mouse genes. *The EMBO Journal*, *21*(23), 6560-6570.

Fraenkel, S., Mostoslavsky, R., Novobrantseva, T. I., Pelanda, R., Chaudhuri, J., Esposito, G., ... Bergman, Y. (2007). Allelic 'choice' governs somatic hypermutation in vivo at the immunoglobulin [kappa]-chain locus. *Nature Immunology*, *8*(7), 715.

Franco, M. M., Prickett, A. R., & Oakey, R. J. (2014). The role of CCCTC-binding factor (CTCF) in genomic imprinting, development, and reproduction. *Biology of Reproduction*, *91*(5).

Fundele, R. H., Norris, M. L., Barton, S. C., Fehlau, M., Howlett, S. K., Mills, W. E., & Surani, M. A. (1990). Temporal and spatial selection against parthenogenetic cells during development of fetal chimeras. *Development, 108*(1), 203-211.

Gaztelumendi, N., & Nogués, C. (2014). Chromosome instability in mouse embryonic stem cells. *Scientific Reports*, *4*, 5324.

Gendrel, A. V., Attia, M., Chen, C. J., Diabangouaya, P., Servant, N., Barillot, E., & Heard, E. (2014). Developmental dynamics and disease potential of random monoallelic gene expression. *Developmental Cell*, *28*(4), 366-380.

Gimelbrant, A., Hutchinson, J. N., Thompson, B. R., & Chess, A. (2007). Widespread monoallelic expression on human autosomes. *Science, 318*(5853), 1136-1140.

Godfrey, P. A., Malnic, B., & Buck, L. B. (2004). The mouse olfactory receptor gene family. *Proceedings of the National Academy of Sciences*, *101*(7), 2156-2161.

Goldmit, M., Ji, Y., Skok, J., Roldan, E., Jung, S., Howard, C., & Bergman, Y. (2005). Epigenetic ontogeny of the Igk locus during B cell development. *Nature Immunology*, *6*(2), 198.

Goodrich, L., Panning, B., & Leung, K. N. (2016). Activators and repressors: A balancing act for X-inactivation. *Seminars in Cell & Developmental Biology, 56*, 3-8.

Gregg, C., Zhang, J., Butler, J. E., Haig, D., & Dulac, C. (2010). Sex-specific parent-of-origin allelic expression in the mouse brain. *Science, 329*(5992), 682-685.

Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., . . . Xu, Y. (2015). Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature, 517*(7536), 640-644.

Haig, D. (2014). Coadaptation and conflict, misconception and muddle, in the evolution of genomic imprinting. *Heredity*, *113*(2), 96.

Haig, D., & Graham, C. (1991). Genomic imprinting and the strange case of the insulin-like growth factor II receptor. *Cell, 64*(6), 1045-1046.

Haig, D., & Westoby, M. (1989). Parent-specific gene expression and the triploid endosperm. *The American Naturalist*, *134*(1), 147-155.

Hark, A. T., Schoenherr, C. J., Katz, D. J., Ingram, R. S., Levorse, J. M., & Tilghman, S. M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature, 405*(6785), 486-489.

Hasegawa, S., Kumagai, M., Hagihara, M., Nishimaru, H., Hirano, K., Kaneko, R., ... Watanabe, M. (2016). Distinct and cooperative functions for the protocadherin-

α, -β and-γ clusters in neuronal survival and axon targeting. *Frontiers in Molecular Neuroscience*, *9*, 155

Hermann, A., Goyal, R., & Jeltsch, A. (2004). The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *Journal of Biological Chemistry*, *279*(46), 48350-48359.

Hikichi, T., Kohda, T., Kaneko-Ishino, T., & Ishino, F. (2003). Imprinting regulation of the murine Meg1/Grb10 and human GRB10 genes; roles of brain-specific promoters and mouse-specific CTCF-binding sites. *Nucleic Acids Research*, *31*(5), 1398-1406.

Hollander, G. A., Zuklys, S., Morel, C., Mizoguchi, E., Mobisson, K., Simpson, S., . . . Burakoff, S. J. (1998). Monoallelic expression of the interleukin-2 locus. *Science, 279*(5359), 2118-2121.

Hutter, B., Helms, V., & Paulsen, M. (2006). Tandem repeats in the CpG islands of imprinted genes. *Genomics, 88*(3), 323-332.

Ishii, T., Serizawa, S., Kohda, A., Nakatani, H., Shiroishi, T., Okumura, K., ... Sakano, H. (2001). Monoallelic expression of the odourant receptor gene and axonal projection of olfactory sensory neurones. *Genes to Cells*, *6*(1), 71-78.

Jeffries, A. R., Perfect, L. W., Ledderose, J., Schalkwyk, L. C., Bray, N. J., Mill, J., & Price, J. (2012). Stochastic choice of allelic expression in human neural stem cells. *Stem Cells*, *30*(9), 1938-1947.

Jenuwein, T., & Allis, C. D. (2001). Translating the histone code. *Science, 293*(5532), 1074-1080.

Jeon, Y., & Lee, J. T. (2011). YY1 tethers Xist RNA to the inactive X nucleation center. *Cell, 146*(1), 119-133.

Ji, Y., Resch, W., Corbett, E., Yamane, A., Casellas, R., & Schatz, D. G. (2010). The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. *Cell, 141*(3), 419-431.

Ji, Y., Zhang, J., Lee, A. I., Cedar, H., & Bergman, Y. (2003). A multistep mechanism for the activation of rearrangement in the immune system. *Proceedings of the National Academy of Sciences*, *100*(13), 7557-7562.

Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A., & Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature, 449*(7159), 248-251.

Johnson, D. R. (1974). Further observations on the hairpin-tail (T hp) mutation in the mouse. *Genetics Research*, *24*(2), 207-214.

Jonkers, I., Barakat, T. S., Achame, E. M., Monkhorst, K., Kenter, A., Rentmeester, E., . . . Gribnau, J. (2009). RNF12 is an X-Encoded dose-dependent activator of X chromosome inactivation. *Cell, 139*(5), 999-1011.

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., & Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature, 429*(6994), 900-903.

Kaneko, R., Kato, H., Kawamura, Y., Esumi, S., Hirayama, T., Hirabayashi, T., & Yagi, T. (2006). Allelic gene regulation of Pcdh-α and Pcdh-γ clusters involving both monoallelic and biallelic expression in single Purkinje cells. *Journal of Biological Chemistry*, *281*(41), 30551-30560.

Kawaguchi, M., Toyama, T., Kaneko, R., Hirayama, T., Kawamura, Y., & Yagi, T. (2008). Relationship between DNA methylation states and transcription of individual isoforms encoded by the protocadherin-α gene cluster. *Journal of Biological Chemistry*, *283*(18), 12064-12075.

Kehayova, P., Monahan, K., Chen, W., & Maniatis, T. (2011). Regulatory elements required for the activation and repression of the protocadherin-α gene cluster. *Proceedings of the National Academy of Sciences*, *108*(41), 17195-17200.

Kelly, B. L., & Locksley, R. M. (2000). Coordinate regulation of the IL-4, IL-13, and IL-5 cytokine cluster in Th2 clones revealed by allelic expression patterns. *The Journal of Immunology*, *165*(6), 2982-2986.

Kim, J. K., Kolodziejczyk, A. A., Ilicic, T., Teichmann, S. A., & Marioni, J. C. (2015). Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. *Nature Communications*, *6*, 8687.

Kim, Y., Lee, H. M., Xiong, Y., Sciaky, N., Hulbert, S. W., Cao, X., . . . Jiang, Y. H. (2017). Targeting the histone methyltransferase G9a activates imprinted genes and improves survival of a mouse model of Prader-Willi syndrome. *Nature Medicine, 23*(2), 213-222.

Kobayashi, S., Isotani, A., Mise, N., Yamamoto, M., Fujihara, Y., Kaseda, K., . . . Okabe, M. (2006). Comparison of gene expression in male and female mouse blastocysts revealed imprinting of the X-linked gene, Rhox5/Pem, at preimplantation stages. *Current Biology, 16*(2), 166-172.

Kratzer, P. G., & Gartler, S. M. (1978). HGPRT activity changes in preimplantation mouse embryos. *Nature, 274*(5670), 503-504.

Krechowec, S. O., Burton, K. L., Newlaczyl, A. U., Nunn, N., Vlatkovic, N., & Plagge, A. (2012). Postnatal changes in the expression pattern of the imprinted signalling protein XLalphas underlie the changing phenotype of deficient mice. *PLoS One, 7*(1), e29753.

Kurukuti, S., Tiwari, V. K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., ... & Ohlsson, R. (2006). CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proceedings of the National Academy of Sciences*, *103*(28), 10684-10689.

Lee, B. K., & Iyer, V. R. (2012). Genome-wide studies of CCCTC-binding factor (CTCF) and cohesin provide insight into chromatin structure and regulation. *Journal of Biological Chemistry*, *287*(37), 30906-30913.

Lee, J. T., Davidow, L. S., & Warshawsky, D. (1999). Tsix, a gene antisense to Xist at the X-inactivation centre. *Nature Genetics, 21*(4), 400-404.

Lee, J. T., & Lu, N. (1999). Targeted mutagenesis of Tsix leads to nonrandom X inactivation. *Cell, 99*(1), 47-57.

Lefebvre, J. L., Kostadinov, D., Chen, W. V., Maniatis, T., & Sanes, J. R. (2012). Protocadherins mediate dendritic self-avoidance in the mammalian nervous system. *Nature, 488*(7412), 517-521.

Li, E., Beard, C., & Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. *Nature, 366*(6453), 362-365.

Li, J. Y., Chai, B. X., Zhang, W., Wang, H., & Mulholland, M. W. (2010). Expression of ankyrin repeat and suppressor of cytokine signaling box protein 4 (Asb-4) in proopiomelanocortin neurons of the arcuate nucleus of mice produces a hyperphagic, lean phenotype. *Endocrinology, 151*(1), 134-142.

Li, J. Y., Kuick, R., Thompson, R. C., Misek, D. E., Lai, Y. M., Liu, Y. Q., ... Gantz, I. (2005). Arcuate nucleus transcriptome profiling identifies ankyrin repeat and suppressor of cytokine signalling box-containing protein 4 as a gene regulated by fasting in central nervous system feeding circuits. *Journal of Neuroendocrinology*, *17*(6), 394-404.

Li, J. Y., Lees-Murdock, D. J., Xu, G. L., & Walsh, C. P. (2004). Timing of establishment of paternal methylation imprints in the mouse. *Genomics, 84*(6), 952-960.

Li, S. M., Valo, Z., Wang, J., Gao, H., Bowers, C. W., & Singer-Sam, J. (2012). Transcriptome-wide survey of mouse CNS-derived cells reveals monoallelic expression within novel gene families. *PLoS One, 7*(2), e31751.

Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P., & Ferguson-Smith, A. C. (2008). A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. *Developmental Cell*, *15*(4), 547-557.

Lingenfelter, P. A., Adler, D. A., Poslinski, D., Thomas, S., Elliott, R. W., Chapman, V. M., & Disteche, C. M. (1998). Escape from X inactivation of Smcx is preceded by silencing during mouse development. *Nature Genetics, 18*(3), 212-213.

Lobanenkov, V. V., Nicolas, R. H., Plumb, M. A., Wright, C. A., & Goodwin, G. H. (1986). Sequence-specific DNA-binding proteins which interact with (G+C)-rich sequences flanking the chicken c-myc gene. *The FEBS Journal*, *159*(1), 181-188.

Lomvardas, S., Barnea, G., Pisapia, D. J., Mendelsohn, M., Kirkland, J., & Axel, R. (2006). Interchromosomal interactions and olfactory receptor choice. *Cell, 126*(2), 403-413.

Lubinsky, M., Herrmann, J., Kosseff, A. L., & Opitz, J. M. (1974). Letter: Autosomal-dominant sex-dependent transmission of the Wiedemann-Beckwith syndrome. *Lancet, 1*(7863), 932.

Lucito, R., Healy, J., Alexander, J., Reiner, A., Esposito, D., Chi, M., ... West, J. A. (2003). Representational oligonucleotide microarray analysis: A high-resolution method to detect genome copy number variation. *Genome Research*, *13*(10), 2291-2305.

Luedi, P. P., Hartemink, A. J., & Jirtle, R. L. (2005). Genome-wide prediction of imprinted murine genes. *Genome Research*, *15*(6), 875-884.

Lyon, M. F. (1961). Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature, 190*, 372-373.

Lyons, D. B., Allen, W. E., Goh, T., Tsai, L., Barnea, G., & Lomvardas, S. (2013). An epigenetic trap stabilizes singular olfactory receptor expression. *Cell, 154*(2), 325-336.

Mager, J., Montgomery, N. D., de Villena, F. P., & Magnuson, T. (2003). Genome imprinting regulated by the mouse Polycomb group protein Eed. *Nature Genetics, 33*(4), 502-507.

Magklara, A., Yen, A., Colquitt, B. M., Clowney, E. J., Allen, W., Markenscoff-Papadimitriou, E., . . . Lomvardas, S. (2011). An epigenetic signature for monoallelic olfactory receptor expression. *Cell, 145*(4), 555-570.

Maher, E. R., & Reik, W. (2000). Beckwith-Wiedemann syndrome: Imprinting in clusters revisited. *Journal of Clinical Investigation*, *105*(3), 247.

Maitra, A., Arking, D. E., Shivapurkar, N., Ikeda, M., Stastny, V., Kassauei, K., . . . Chakravarti, A. (2005). Genomic alterations in cultured human embryonic stem cells. *Nature Genetics, 37*(10), 1099-1103.

Marahrens, Y., Panning, B., Dausman, J., Strauss, W., & Jaenisch, R. (1997). Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes & Development*, *11*(2), 156-166.

Marinov, G. K., Williams, B. A., McCue, K., Schroth, G. P., Gertz, J., Myers, R. M., & Wold, B. J. (2014). From single-cell to cell-pool transcriptomes: Stochasticity in gene expression and RNA splicing. *Genome Research, 24*(3), 496-510.

McGrath, J., & Solter, D. (1983). Nuclear transplantation in mouse embryos. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, *228*(2), 355-362.

McGrath, J., & Solter, D. (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell, 37*(1), 179-183.

McHugh, C. A., Chen, C. K., Chow, A., Surka, C. F., Tran, C., McDonel, P., . . . Guttman, M. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature, 521*(7551), 232-236.

McLaren, A. (2003). Primordial germ cells in the mouse. *Developmental Biology, 262*(1), 1-15.

Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L., & Bird, A. P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell, 58*(3), 499-507.

Messerschmidt, D. M., de Vries, W., Ito, M., Solter, D., Ferguson-Smith, A., & Knowles, B. B. (2012). Trim28 is required for epigenetic stability during mouse oocyte to embryo transition. *Science, 335*(6075), 1499-1502.

Miller, C. A., & Sweatt, J. D. (2007). Covalent modification of DNA regulates memory formation. *Neuron, 53*(6), 857-869.

Monahan, K., Rudnick, N. D., Kehayova, P. D., Pauli, F., Newberry, K. M., Myers, R. M., & Maniatis, T. (2012). Role of CCCTC binding factor (CTCF) and cohesin in the generation of single-cell diversity of protocadherin-α gene expression. *Proceedings of the National Academy of Sciences*, *109*(23), 9125-9130.

Monfort, A., Di Minin, G., Postlmayr, A., Freimann, R., Arieti, F., Thore, S., & Wutz, A. (2015). Identification of Spen as a crucial factor for Xist function through forward genetic screening in haploid embryonic stem cells. *Cell Reports*, *12*(4), 554-561.

Monkhorst, K., Jonkers, I., Rentmeester, E., Grosveld, F., & Gribnau, J. (2008). X inactivation counting and choice is a stochastic process: Evidence for involvement of an X-linked activator. *Cell, 132*(3), 410-421.

Mostoslavsky, R., Singh, N., Tenzen, T., Goldmit, M., Gabay, C., Elizur, S., . . . Bergman, Y. (2001). Asynchronous replication and allelic exclusion in the immune system. *Nature, 414*(6860), 221-225.

Murrell, A., Heeson, S., & Reik, W. (2004). Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. *Nature Genetics, 36*(8), 889-893.

Nag, A., Savova, V., Fung, H. L., Miron, A., Yuan, G. C., Zhang, K., & Gimelbrant, A. A. (2013). Chromatin signature of widespread monoallelic expression. *Elife, 2*, e01256.

Nag, A., Vigneau, S., Savova, V., Zwemer, L. M., & Gimelbrant, A. A. (2015). Chromatin signature identifies monoallelic gene expression across mammalian cell types. *G3 (Bethesda), 5*(8), 1713-1720.

Nagano, T., Mitchell, J. A., Sanz, L. A., Pauler, F. M., Ferguson-Smith, A. C., Feil, R., & Fraser, P. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science, 322*(5908), 1717-1720.

Nakase, H., Takahama, Y., & Akamatsu, Y. (2003). Effect of CpG methylation on RAG1/RAG2 reactivity: Implications of direct and indirect mechanisms for controlling V (D) J cleavage. *EMBO Reports*, *4*(8), 774-780.

Nguyen, D. K., & Disteche, C. M. (2006). Dosage compensation of the active X chromosome in mammals. *Nature Genetics, 38*(1), 47-53.

Nicholls, R. D., Knoll, J. H., Butler, M. G., Karam, S., & Lalande, M. (1989). Genetic imprinting suggested by maternal heterodisomy in nondeletion Prader-Willi syndrome. *Nature, 342*(6247), 281-285.

Ohno, S., Kaplan, W. D., & Kinosita, R. (1959). Formation of the sex chromatin by a single X-chromosome in liver cells of Rattus norvegicus. *Experimental Cell Research*, *18*(2), 415-418.

Okamoto, I., Otte, A. P., Allis, C. D., Reinberg, D., & Heard, E. (2004). Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science, 303*(5658), 644-649.

Ooi, S. K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., . . . Bestor, T. H. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature, 448*(7154), 714-717.

Pant, V., Kurukuti, S., Pugacheva, E., Shamsuddin, S., Mariano, P., Renkawitz, R., ... Ohlsson, R. (2004). Mutation of a single CTCF target site within the H19 imprinting control region leads to loss of Igf2 imprinting and complex patterns of de novo methylation upon maternal inheritance. *Molecular and Cellular Biology*, *24*(8), 3497-3504.

Patil, D. P., Chen, C. K., Pickering, B. F., Chow, A., Jackson, C., Guttman, M., & Jaffrey, S. R. (2016). m(6)A RNA methylation promotes XIST-mediated transcriptional repression. *Nature, 537*(7620), 369-373.

Paulsen, M., El-Maarri, O., Engemann, S., Strödicke, M., Franck, O., Davies, K., ... Walter, J. (2000). Sequence conservation and variability of imprinting in the Beckwith–Wiedemann syndrome gene cluster in human and mouse. *Human Molecular Genetics*, *9*(12), 1829-1841.

Pelanda, R. (2014). Dual immunoglobulin light chain B cells: Trojan horses of autoimmunity? *Current Opinion in Immunology*, *27*, 53-59.

Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S., & Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. *Nature, 379*(6561), 131-137.

Pereira, J. P., Girard, R., Chaby, R., Cumano, A., & Vieira, P. (2003). Monoallelic expression of the murine gene encoding Toll-like receptor 4. *Nature Immunology*, *4*(5), 464.

Perez, J. D., Rubinstein, N. D., & Dulac, C. (2016). New perspectives on genomic imprinting, an essential and multifaceted mode of epigenetic control in the developing and adult brain. *Annual Review of Neuroscience*, *39*, 347-384.

Pernis, B., Chiappino, G., Kelus, A. S., & Gell, P. G. (1965). Cellular localization of immunoglobulins with different allotypic specificities in rabbit lymphoid tissues. *Journal of Experimental Medicine*, *122*(5), 853-876.

Perry, J., Palmer, S., Gabriel, A., & Ashworth, A. (2001). A short pseudoautosomal region in laboratory mice. *Genome Research, 11*(11), 1826-1832.

Peters, J. (2014). The role of genomic imprinting in biology and disease: An expanding view. *Nature Reviews. Genetics*, *15*(8), 517.

Peters, J., Wroe, S. F., Wells, C. A., Miller, H. J., Bodle, D., Beechey, C. V., ... Kelsey, G. (1999). A cluster of oppositely imprinted transcripts at the Gnas locus in the distal imprinting region of mouse chromosome 2. *Proceedings of the National Academy of Sciences*, *96*(7), 3830-3835.

Pinter, S. F., Colognori, D., Beliveau, B. J., Sadreyev, R. I., Payer, B., Yildirim, E., ... Lee, J. T. (2015). Allelic imbalance is a prevalent and tissue-specific feature of the mouse transcriptome. *Genetics*, *200*(2), 537-549.

Plagge, A., Gordon, E., Dean, W., Boiani, R., Cinti, S., Peters, J., & Kelsey, G. (2004). The imprinted signaling protein XL alpha s is required for postnatal adaptation to feeding. *Nature Genetics, 36*(8), 818-826.

Prickett, A. R., Barkas, N., McCole, R. B., Hughes, S., Amante, S. M., Schulz, R., & Oakey, R. J. (2013). Genome-wide and parental allele-specific analysis of CTCF and cohesin DNA binding in mouse brain reveals a tissue-specific binding pattern and an association with imprinted differentially methylated regions. *Genome Research, 23*(10), 1624-1635.

Prickett, A. R., & Oakey, R. J. (2012). A survey of tissue-specific genomic imprinting in mammals. *Molecular Genetics and Genomics*, *287*(8), 621-630.

Priest, J. H., Heady, J. E., & Priest, R. E. (1967). Delayed onset of replication of human X chromosomes. *The Journal of Cell Biology*, *35*(2), 483.

Qian, N., Frank, D., O'Keefe, D., Dao, D., Zhao, L., Yuan, L., ... Tycko, B. (1997). The IPL gene on chromosome 11p15. 5 is imprinted in humans and mice and is similar to TDAG51, implicated in Fas expression and apoptosis. *Human Molecular Genetics*, *6*(12), 2021-2029.

Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., ... Trono, D. (2011). In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Molecular Cell*, *44*(3), 361-372.

Raefski, A. S., & O'Neill, M. J. (2005). Identification of a cluster of X-linked imprinted genes in mice. *Nature Genetics*, *37*(6), 620.

Rastan, S., & Brown, S. D. M. (1990). The search for the mouse X-chromosome inactivation centre. *Genetics Research*, *56*(2-3), 99-106.

Ratel, D., Ravanat, J. L., Berger, F., & Wion, D. (2006). N6-methyladenine: The other methylated base of DNA. *Bioessays, 28*(3), 309-315.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., . . . Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature, 406*(6796), 593-599.

Reik, W., & Walter, J. (2001). Genomic imprinting: Parental influence on the genome. *Nature Reviews. Genetics*, *2*(1), 21.

Reinius, B., Mold, J. E., Ramskold, D., Deng, Q., Johnsson, P., Michaelsson, J., . . . Sandberg, R. (2016). Analysis of allelic expression patterns in clonal somatic cells by single-cell RNA-seq. *Nature Genetics, 48*(11), 1430-1435.

Reinius, B., & Sandberg, R. (2015). Random monoallelic expression of autosomal genes: Stochastic transcription and allele-level regulation. *Nature Reviews. Genetics*, *16*(11), 653.

Ribich, S., Tasic, B., & Maniatis, T. (2006). Identification of long-range regulatory elements in the protocadherin-α gene cluster. *Proceedings of the National Academy of Sciences*, *103*(52), 19719-19724.

Rinn, J. L., & Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. *Annual Review of Biochemistry*, *81*, 145-166.

Sanchez-Delgado, M., Riccio, A., Eggermann, T., Maher, E. R., Lapunzina, P., Mackay, D., & Monk, D. (2016). Causes and consequences of multi-locus imprinting disturbances in humans. *Trends in Genetics*, *32*(7), 444-455.

Santoro, F., and Barlow, D.P. (2011). Developmental control of imprinted expression by macro non-coding RNAs. *Seminars in Cell & Developmental Biology 22*, 328-335.

Saraiva, L. R., Ibarra-Soria, X., Khan, M., Omura, M., Scialdone, A., Mombaerts, P., ... Logan, D. W. (2015). Hierarchical deconstruction of mouse olfactory sensory neurons: from whole mucosa to single-cell RNA-seq. *Scientific Reports*, *5*, 18178.

Sarma, K., Cifuentes-Rojas, C., Ergun, A., Del Rosario, A., Jeon, Y., White, F., . . . Lee, J. T. (2014). ATRX directs binding of PRC2 to Xist RNA and Polycomb targets. *Cell, 159*(4), 869-883.

Savova, V., Chun, S., Sohail, M., McCole, R. B., Witwicki, R., Gai, L., . . . Gimelbrant, A. A. (2016). Genes with monoallelic expression contribute disproportionately to genetic diversity in humans. *Nature Genetics, 48*(3), 231-237.

Savova, V., Patsenker, J., Vigneau, S., & Gimelbrant, A. A. (2016). dbMAE: The database of autosomal monoallelic expression. *Nucleic Acids Research*, *44*(D1), D753-D756.

Savova, V., Vinogradova, S., Pruss, D., Gimelbrant, A. A., & Weiss, L. A. (2017). Risk alleles of genes with monoallelic expression are enriched in gain-of-function variants and depleted in loss-of-function variants for neurodevelopmental disorders. *Molecular Psychiatry*, mp201713.

Schaller, F., Watrin, F., Sturny, R., Massacrier, A., Szepetowski, P., & Muscatelli, F. (2010). A single postnatal injection of oxytocin rescues the lethal feeding behaviour in mouse newborns deficient for the imprinted Magel2 gene. *Human Molecular Genetics*, *19*(24), 4895-4905.

Searle, A. G., & Beechey, C. V. (1978). Complementation studies with mouse translocations. *Cytogenetic and Genome Research*, *20*(1-6), 282-303.

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., . . . Reik, W. (2012). The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Molecular Cell, 48*(6), 849-862.

Serizawa, S., Miyamichi, K., & Sakano, H. (2004). One neuron–one receptor rule in the mouse olfactory system. *Trends in Genetics*, *20*(12), 648-653.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., . . . Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell, 119*(7), 941-953.

Simon, I., Tenzen, T., Reubinoff, B. E., Hillman, D., McCarrey, J. R., & Cedar, H. (1999). Asynchronous replication of imprinted genes is established in the gametes and maintained during development. *Nature, 401*(6756), 929-932.

Sims, R. J., 3rd, Nishioka, K., & Reinberg, D. (2003). Histone lysine methylation: A signature for chromatin function. *Trends in Genetics, 19*(11), 629-639.

Singh, N., Ebrahimi, F. A., Gimelbrant, A. A., Ensminger, A. W., Tackett, M. R., Qi, P., . . . Chess, A. (2003). Coordination of the random asynchronous replication of autosomal loci. *Nature Genetics, 33*(3), 339-341.

Sleutels, F., Zwart, R., & Barlow, D. P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature, 415*(6873), 810-813.

Smallwood, S. A., & Kelsey, G. (2012). De novo DNA methylation: A germ cell perspective. *Trends in Genetics, 28*(1), 33-42.

Smilinich, N. J., Day, C. D., Fitzpatrick, G. V., Caldwell, G. M., Lossie, A. C., Cooper, P. R., ... Nicholls, R. D. (1999). A maternally methylated CpG island in KvLQT1 is associated with an antisense paternal transcript and loss of imprinting in Beckwith–Wiedemann syndrome. *Proceedings of the National Academy of Sciences*, *96*(14), 8064-8069.

Soellner, L., Begemann, M., Mackay, D. J., Grønskov, K., Tümer, Z., Maher, E. R., ... Netchine, I. (2017). Recent advances in imprinting disorders. *Clinical Genetics*, *91*(1), 3-13.

Sonoda, E., Pewzner-Jung, Y., Schwers, S., Taki, S., Jung, S., Eilat, D., & Rajewsky, K. (1997). B cell development under the condition of allelic inclusion. *Immunity, 6*(3), 225-233.

Spence, J. E., Perciaccante, R. G., Greig, G. M., Willard, H. F., Ledbetter, D. H., Hejtmancik, J. F., . . . Beaudet, A. L. (1988). Uniparental disomy as a mechanism for human genetic disease. *American Journal of Human Genetics, 42*(2), 217-226.

Spencer, H. G., & Clark, A. G. (2014). Non-conflict theories for the evolution of genomic imprinting. *Heredity*, *113*(2), 112.

Splinter, E., Heath, H., Kooren, J., Palstra, R. J., Klous, P., Grosveld, F., ... de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus. *Genes & Development*, *20*(17), 2349-2354.

Stoger, R., Kubicka, P., Liu, C. G., Kafri, T., Razin, A., Cedar, H., & Barlow, D. P. (1993). Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal. *Cell, 73*(1), 61-71.

Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H., & Tajima, S. (2004). DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *Journal of Biological Chemistry*, *279*(26), 27816-27823.

Surani, M. A., & Barton, S. C. (1983). Development of gynogenetic eggs in the mouse: Implications for parthenogenetic embryos. *Science, 222*(4627), 1034-1036.

Surani, M. A., Barton, S. C., & Norris, M. L. (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature, 308*(5959), 548-550.

Takizawa, T., Gudla, P. R., Guo, L., Lockett, S., & Misteli, T. (2008). Allele-specific nuclear positioning of the monoallelically expressed astrocyte marker GFAP. *Genes & Development*, *22*(4), 489-498.

Tasic, B., Nabholz, C. E., Baldwin, K. K., Kim, Y., Rueckert, E. H., Ribich, S. A., . . . Maniatis, T. (2002). Promoter choice determines splice site selection in protocadherin alpha and gamma pre-mRNA splicing. *Molecular Cell, 10*(1), 21-33.

Tian, D., Sun, S., & Lee, J. T. (2010). The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell, 143*(3), 390-403.

Tsuchiya, K. D., Greally, J. M., Yi, Y., Noel, K. P., Truong, J. P., & Disteche, C. M. (2004). Comparative sequence and x-inactivation analyses of a domain of escape in human xp11.2 and the conserved segment in mouse. *Genome Research, 14*(7), 1275-1284.

Vettermann, C., & Schlissel, M. S. (2010). Allelic exclusion of immunoglobulin genes: Models and mechanisms. *Immunological Reviews*, *237*(1), 22-42.

Vilkaitis, G., Suetake, I., Klimašauskas, S., & Tajima, S. (2005). Processive methylation of hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase. *Journal of Biological Chemistry*, *280*(1), 64-72.

Wan, L. B., & Bartolomei, M. S. (2008). Regulation of imprinting in clusters: Noncoding RNAs versus insulators. *Advances in Genetics*, *61*, 207-223.

Wang, H., Maurano, M. T., Qu, H., Varley, K. E., Gertz, J., Pauli, F., ... Thurman, R. E. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Research*, *22*(9), 1680-1688.

Wang, L., Brown, J. L., Cao, R., Zhang, Y., Kassis, J. A., & Jones, R. S. (2004). Hierarchical recruitment of polycomb group silencing complexes. *Molecular Cell, 14*(5), 637-646.

Wang, X., Soloway, P. D., & Clark, A. G. (2010). Paternally biased X inactivation in mouse neonatal brain. *Genome Biology*, *11*(7), R79.

Wang, X., Weiner, J. A., Levi, S., Craig, A. M., Bradley, A., & Sanes, J. R. (2002). Gamma protocadherins are required for survival of spinal interneurons. *Neuron, 36*(5), 843-854.

Wang, Z., Willard, H. F., Mukherjee, S., & Furey, T. S. (2006). Evidence of influence of genomic DNA sequence on human X chromosome inactivation. *PLoS Computational Biology*, *2*(9), e113.

Wilkins, J. F., & Haig, D. (2003). What good is genomic imprinting: The function of parent-specific gene expression. *Nature Reviews. Genetics*, *4*(5), 359.

Wolf, J. B., & Hager, R. (2006). A maternal–offspring coadaptation theory for the e volution of genomic imprinting. *PLoS Biology*, *4*(12), e380.

Wu, H., Luo, J., Yu, H., Rattner, A., Mo, A., Wang, Y., . . . Nathans, J. (2014). Cellular resolution maps of X chromosome inactivation: Implications for neural development, function, and disease. *Neuron, 81*(1), 103-119.

Wu, H., Yang, L., & Chen, L. L. (2017). The diversity of long noncoding RNAs and their generation. *Trends in Genetics, 33*, 540-552.

Wu, Q., & Maniatis, T. (1999). A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell, 97*(6), 779-790.

Wu, Q., Zhang, T., Cheng, J. F., Kim, Y., Grimwood, J., Schmutz, J., ... Maniatis, T. (2001). Comparative DNA sequence analysis of mouse and human protocadherin gene clusters. *Genome Research*, *11*(3), 389-404.

Wutz, A., Smrzka, O. W., Schweifer, N., Schellander, K., Wagner, E. F., & Barlow, D. P. (1997). Imprinted expression of the Igf2r gene depends on an intronic CpG island. *Nature, 389*(6652), 745-749.

Xin, Z., Allis, C. D., & Wagstaff, J. (2001). Parent-specific complementary patterns of histone H3 lysine 9 and H3 lysine 4 methylation at the Prader-Willi syndrome imprinting center. *The American Journal of Human Genetics*, *69*(6), 1389-1394.

Xu, J., Carter, A. C., Gendrel, A. V., Attia, M., Loftus, J., Greenleaf, W. J., . . . Chang, H. Y. (2017). Landscape of monoallelic DNA accessibility in mouse embryonic stem cells and neural progenitor cells. *Nature Genetics, 49*(3), 377-386.

Yagi, T. (2008). Clustered protocadherin family. *Development, Growth & Differentiation*, *50*(s1).

Yagi, T. (2012). Molecular codes for neuronal individuality and cell assembly in the brain. *Frontiers in Molecular Neuroscience*, *5*, 45.

Yamaguchi, S., Shen, L., Liu, Y., Sendler, D., & Zhang, Y. (2013). Role of Tet1 in erasure of genomic imprinting. *Nature, 504*(7480), 460-464.

Yang, F., Babak, T., Shendure, J., & Disteche, C. M. (2010). Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Research, 20*(5), 614-622.

Yoder, J. A., Soman, N. S., Verdine, G. L., & Bestor, T. H. (1997). DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. *Journal of Molecular Biology*, *270*(3), 385-395.

Yokota, S., Hirayama, T., Hirano, K., Kaneko, R., Toyoda, S., Kawamura, Y., ... Yagi, T. (2011). Identification of the cluster control region for the protocadherin-β genes located beyond the protocadherin-γ cluster. *Journal of Biological Chemistry*, *286*(36), 31885-31895.

Yoon, Y. S., Jeong, S., Rong, Q., Park, K. Y., Chung, J. H., & Pfeifer, K. (2007). Analysis of the H19ICR insulator. *Molecular and Cellular Biology*, *27*(9), 3499-3510.

Zhang, T., Termanis, A., Özkan, B., Bao, X. X., Culley, J., de Lima Alves, F., ... Stancheva, I. (2016). G9a/GLP complex maintains imprinted DNA methylation in embryonic stem cells. *Cell Reports*, *15*(1), 77-85.

Zhang, X., & Firestein, S. (2002). The olfactory receptor gene superfamily of the mouse. *Nature Neuroscience*, *5*(2), 124.

Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J., & Lee, J. T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science, 322*(5902), 750-756.

Zozulya, S., Echeverri, F., & Nguyen, T. (2001). The human olfactory receptor repertoire. *Genome Biology*, *2*(6), research0018-1.

Zwemer, L. M., Zak, A., Thompson, B. R., Kirby, A., Daly, M. J., Chess, A., & Gimelbrant, A. A. (2012). Autosomal monoallelic expression in the mouse. *Genome Biology*, *13*(2), R10.

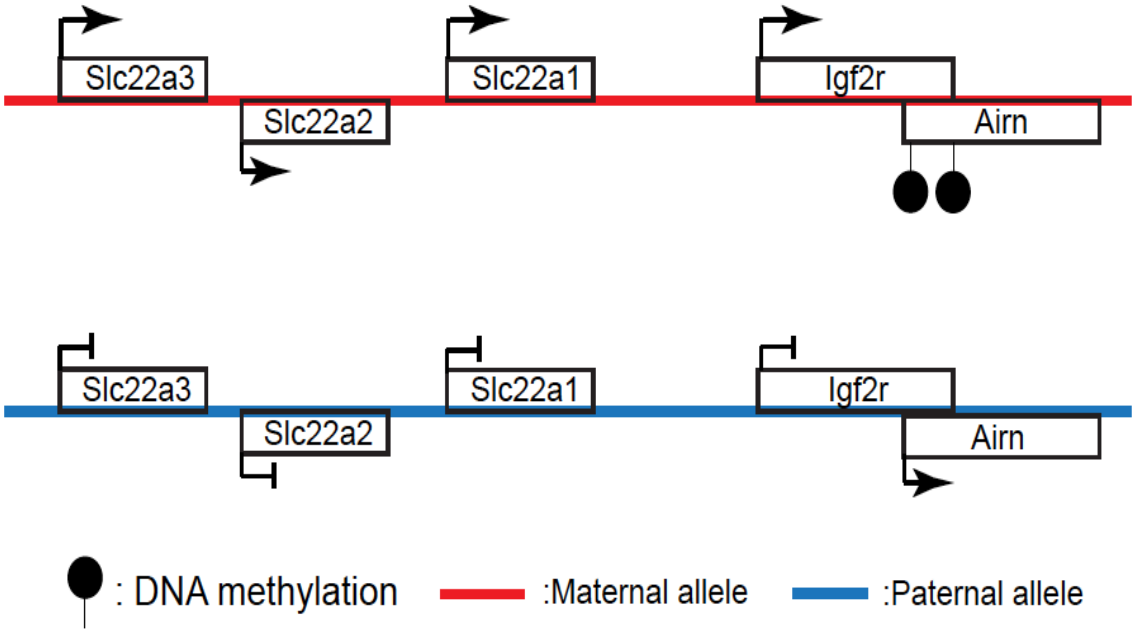


Figure 1.1 lncRNA model at imprinted gene cluster. Airn is a lncRNA, which is antisense to Igf2r in the gene cluster. The Airn promoter is hypermethylated on the maternal allele but not on the paternal allele. The Airn expressed paternally represses Igf2r expression on the paternal allele by transcriptional interference.

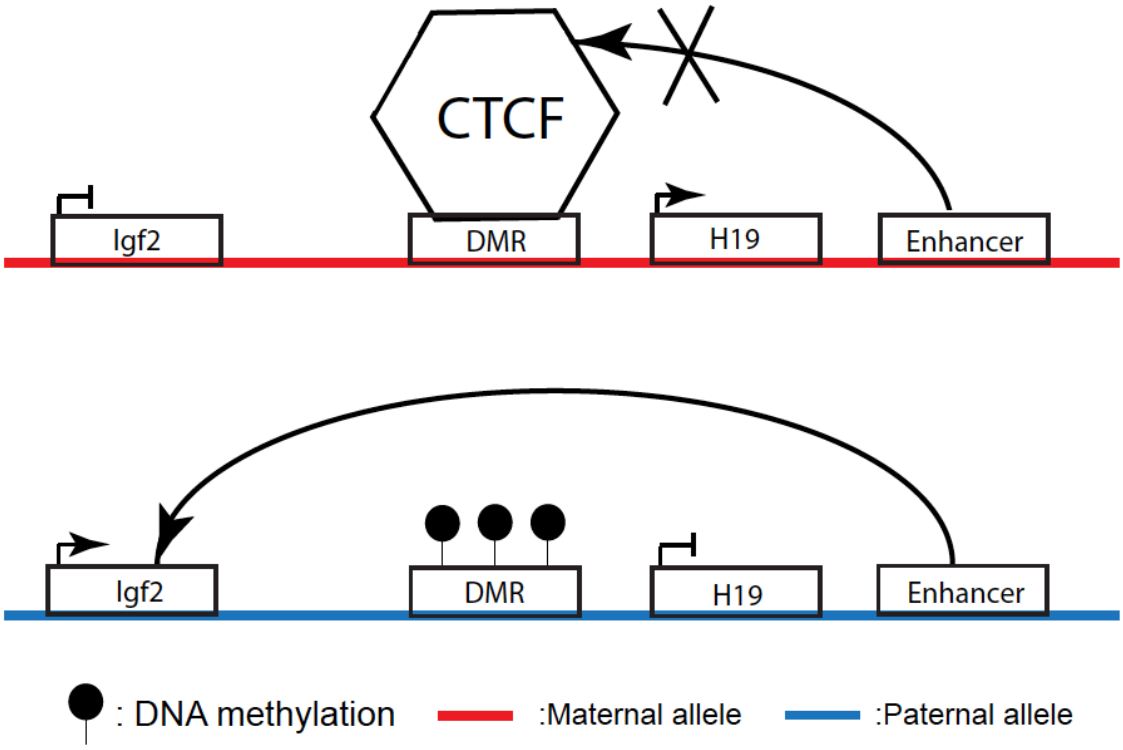


Figure 1.2 Insulator model at imprinted gene cluster. The CTCF is bound to the unmethylated maternal allele, while its binding is blocked on the methylated paternal allele because of the methylation-sensitive feature on the differential methylation region (DMR). Thus, Igf2 expresses on the paternal allele due to an accessible enhancer, while H19 expresses on the maternal allele.

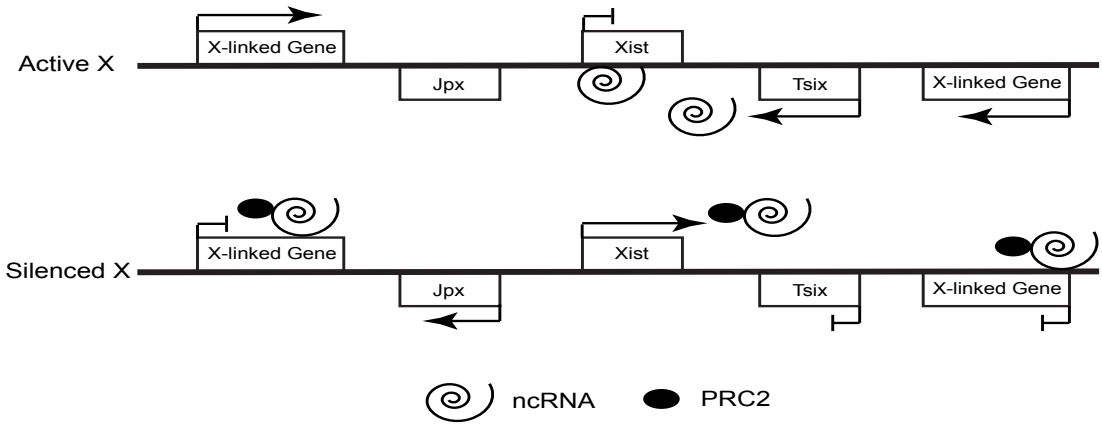


Figure 1.3 A basic model of random X-inactivation. In a silenced X chromosome, Xist is expressed to inactivate X-linked genes' expression, and other factors involving in X-inactivation has been identified (e.g., Jpx and PRC2). In active X chromosome, Tsix is expressed to repress Xist expression.

CHAPTER 2

NONCANONICAL GENOMIC IMPRINTING IN OFFSPRING

The following chapter is reprinted with permission from Cell Press.

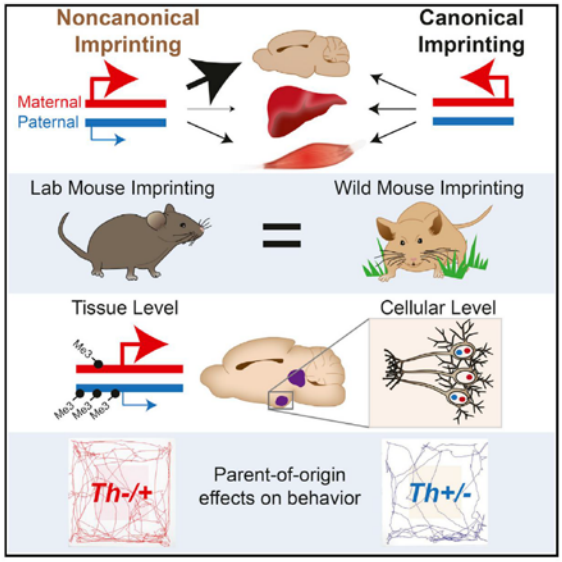
Bonthuis, P. J., Huang, W. C., Hörndli, C. N. S., Ferris, E., Cheng, T., & Gregg, C. (2015). Noncanonical genomic imprinting effects in offspring. *Cell reports*, *12*(6), 979-991.

Article

Cell Reports

Noncanonical Genomic Imprinting Effects in Offspring

Graphical Abstract



Highlights

- Sensitive RNA-seq approach detects tissue-specific maternal and paternal allele biases
- Noncanonical imprinting effects are conserved in wild populations
- Nascent RNA in situ hybridization reveals allelic effects in subpopulations of neurons
- Noncanonical imprinting influences the monoamine pathway and offspring behavior

Authors

Paul J. Bonthuis, Wei-Chao Huang, Cornelia N. Stacher Hörndli, Elliott Ferris, Tong Cheng, Christopher Gregg

Correspondence

chris.gregg@neuro.utah.edu

In Brief

Canonical imprinting involves silencing of the maternal or paternal allele. Bonthuis et al. describe tissue-specific noncanonical imprinting effects involving maternal or paternal allele expression biases. Noncanonical imprinted genes are enriched in the brain and, at the cellular level, exhibit allele-specific expression effects in discrete subpopulations of neurons. They find that noncanonical imprinting can lead to parent-of-origin effects for inherited mutations that impact brain function and behavior.

Accession Numbers

GSE70484

Bonthuis et al., 2015, Cell Reports *12*, 979–991
August 11, 2015 ©2015 The Authors
http://dx.doi.org/10.1016/j.celrep.2015.07.017

CellPress

Cell Reports
Article

OPEN ACCESS
CellPress

Noncanonical Genomic Imprinting Effects in Offspring

Paul J. Bonthuis,[1,3] Wei-Chao Huang,[1,3] Cornelia N. Stacher Hörndli,[1,3] Elliott Ferris,[1,3] Tong Cheng,[1,3] and Christopher Gregg[1,2,*]
[1]Department of Neurobiology & Anatomy
[2]Department of Human Genetics
University of Utah School of Medicine, Salt Lake City, UT 84132-3401, USA
[3]Co-first author
*Correspondence: chris.gregg@neuro.utah.edu
http://dx.doi.org/10.1016/j.celrep.2015.07.017

SUMMARY

Here, we describe an RNA-sequencing (RNA-seq)-based approach that accurately detects even modest maternal or paternal allele expression biases at the tissue level, which we call noncanonical genomic imprinting effects. We profile imprinting in the arcuate nucleus (ARN) and dorsal raphe nucleus of the female mouse brain as well as skeletal muscle (mesodermal) and liver (endodermal). Our study uncovers hundreds of noncanonical autosomal and X-linked imprinting effects. Noncanonical imprinting is highly tissue-specific and enriched in the ARN, but rare in the liver. These effects are reproducible across different genetic backgrounds and associated with allele-specific chromatin. Using in situ hybridization for nascent RNAs, we discover that autosomal noncanonical imprinted genes with a tissue-level allele bias exhibit allele-specific expression effects in subpopulations of neurons in the brain in vivo. We define noncanonical imprinted genes that regulate monoamine signaling and determine that these effects influence the impact of inherited mutations on offspring behavior.

INTRODUCTION

Many inherited genetic risk factors for complex disorders, such as neuropsychiatric disorders, are heterozygous in the affected individuals (Huguet et al., 2013). Therefore, understanding allele-specific expression effects in different tissues and cell types is essential for understanding how inherited mutations may impact offspring. Genomic imprinting is a heritable form of epigenetic gene regulation that results in preferential expression of the maternal or paternal allele for at least 100 genes in mammals (Bartolomei and Ferguson-Smith, 2011). In females, imprinting can influence both autosomal and X-linked genes, and a consequence of imprinting is that the effect of an inherited mutation is influenced by the parental origin.

Canonical imprinting is associated with complete silencing of one gene copy. Indeed, models of the Kinship Theory for the evolution of imprinting predict that evolutionary parental conflicts drive complete silencing of one parent's allele at loci that influence offspring demands on maternal resources (Haig, 2000). However, early studies noted that some imprinted genes exhibit a bias to express either the maternal or paternal allele, rather than complete silencing (Khatib, 2007). Compared to canonical imprinted genes, genes that exhibit allele expression biases might be associated with different mechanisms, functions, and selective pressures. Here, we refer to these effects as "noncanonical imprinting effects." Previously, we devised an approach to profile imprinting in the developing and adult mouse brain using RNA sequencing (RNA-seq) (Gregg et al., 2010a, 2010b) and uncovered noncanonical imprinting effects that influence the expression of hundreds of genes. On the other hand, some other studies of imprinting in somatic tissues found very few novel imprinted genes in mice (Babak et al., 2008; Wang et al., 2008), whereas a study of the mouse liver uncovered 535 imprinted genes (Goncalves et al., 2012). Our findings have been debated (DeVeale et al., 2012), and two recent studies of imprinting in different mouse tissues reached different conclusions regarding the prevalence of imprinting and the identity of the novel imprinted genes detected (Babak et al., 2015; Crowley et al., 2015). Thus, noncanonical imprinting effects in the genome remain poorly understood, and the mechanisms involved and possible function(s) of noncanonical imprinting are unknown.

Here, we devise and apply improved methods to detect imprinting in different tissues by RNA-seq. We perform a genome-wide analysis of canonical and noncanonical imprinting effects in adult female mice for the arcuate nucleus of the hypothalamus (ARN), the dorsal raphe nucleus (DRN) of the midbrain, the liver (endoderm-derived), and skeletal muscle (mesoderm derived). Neuronal circuits in the ARN regulate the endocrine system, feeding, energy expenditure, and blood glucose homeostasis (Gao and Horvath, 2007; Sternson, 2013), while the DRN, a major serotonergic nucleus, influences stress and anxiety, arousal, feeding, reward, social behaviors, and pain (Challis et al., 2013; Dölen et al., 2013; Lowry et al., 2008; Michelsen et al., 2007; Monti, 2010; Wang and Nakai, 1994). By comparing imprinting in the brain to the liver and muscle, we examine the prevalence of canonical and noncanonical imprinting effects in different tissue types. By comparing the ARN and DRN, we determine whether imprinting differs between brain regions with important roles in human health. Our study reveals that

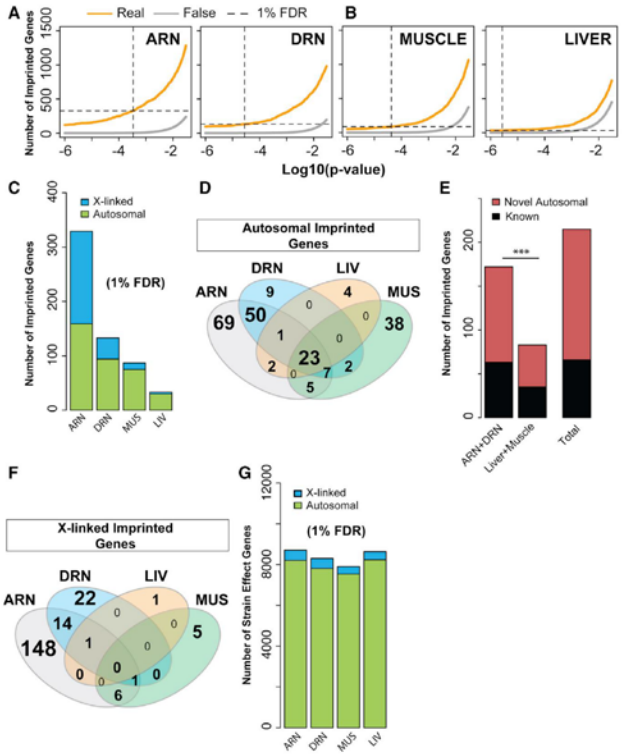


Figure 1. Detection of Imprinting Effects in the Adult Female ARN, DRN, Liver, and Muscle

(A and B) Number of imprinted genes detected by RNA-seq (orange line) and the estimated false-positives (gray line) at different p value cutoffs for the ARN, DRN, muscle, and liver (dashed line is 1% FDR).

(C) Number of autosomal (green bars) and X-linked (blue bars) imprinted genes in each tissue at the 1% FDR.

(D) Venn diagram of autosomal imprinted genes detected in each tissue at the 1% FDR.

(E) Number of novel (maroon bar) and known (black bar) autosomal imprinted genes uncovered in the ARN and DRN (neural) compared to the muscle and liver (non-neural), as well as the total number in all tissues.

(F) Venn diagram of X-lined imprinted genes detected at the 1% FDR cutoff in each tissue.

(G) Number of autosomal and X-linked genes that exhibit genetic strain effects in each tissue at the 1% FDR.

noncanonical imprinting effects are tissue specific and impact a few hundred autosomal and X-linked genes. We perform extensive independent validation studies that support our findings and demonstrate that noncanonical imprinting occurs in wild-derived outbred populations and involves allele-specific chromatin modifications. At the cellular level, noncanonical imprinted genes exhibit allele-specific expression effects in discrete subpopulations of neurons in the brain. These effects influence multiple genes in the monoamine pathway and cause parent-of-origin effects on offspring behavior for inherited heterozygous mutations in *tyrosine hydroxylase* (*Th*). Our results have important implications for understanding the genetic and epigenetic architecture underlying brain function and complex phenotypes.

RESULTS

Discovery of Novel and Tissue-Specific Imprinting Effects in Adult Female Mice

To detect imprinting effects in the ARN, DRN, liver, and muscle, we generated adult female F1 hybrid offspring from reciprocal crosses of CastEiJ (Cast) and C57BL/6J (B6) inbred strains and performed RNA-seq to profile the transcriptome of the initial (F1i: Cast mother × B6 father) and reciprocal (F1r: B6 mother × Cast father) hybrid offspring. We use base calls at SNP sites to distinguish expression from maternal and paternal alleles, as previously described (Gregg et al., 2010a, 2010b). For each tissue, we perform eight to nine biological replicates for each cross and deep sequencing, generating 80–100 million 59-bp single-end reads per replicate (Table S1). We made many advances to improve our methodology for detecting imprinting effects as detailed in the Supplemental Information (Figure S1). With these methods, we analyzed the ARN, DRN, liver, and muscle of adult female mice and determined the number of imprinted genes detected across a range of p value cutoffs ($p = 1 \times 10^{-6}$ to $p = 0.1$) (Figure 1A). The number of false-positives was estimated using a permutation test (Figure S1) and revealed that hundreds of genes exhibit significant imprinting effects in the ARN and DRN (Figure 1A), but fewer exist in the muscle and very few in the liver (Figure 1B). For each data set, we identified the p value cutoff that yields a conservative 1% false discovery rate (FDR) to define a high confidence set of imprinted genes (Figures 1A–1C). We identified 328 imprinted genes in the ARN, of which 158 are autosomal and 170 are X-linked (Figure 1C). We found that the ARN has 79% more autosomal imprinted genes than the DRN (93 imprinted genes), 110% more than the muscle (75 imprinted genes), and over 5-fold more than the liver (30 imprinted genes) (Figures 1C and 1D). Out of the 69 imprinted genes detected specifically in the ARN (Figure 1D), 48 genes (70%) showed the same direction of allele bias in the DRN,

but the magnitude of the bias was stronger in the ARN (ARN mean allele bias, 65%; DRN mean allele bias, 16%).

Our analysis uncovered autosomal imprinting effects that are specific to each tissue type (Figure 1D). We found over twice as many autosomal imprinted genes in the brain (ARN + DRN: 172 genes) compared to the nonneural tissues (muscle + liver: 83 genes), which is a significant difference ($p = 7.5 \times 10^{-5}$, Fisher's exact test) (Figure 1E). At the 1% FDR cutoff, we further detected 198 X-linked genes in total, and 75% of these were specifically identified in the ARN (170 X-linked genes) (Figures 1C and 1F). Thus, both autosomal and X-linked imprinting effects are enriched in the brain, and highly enriched in the ARN.

We compiled a list of 151 accepted imprinted genes from available public repositories (Schulz et al., 2008) and found that 98 are ensembl-annotated for the mouse. From this list, we determined that 142 of the 209 autosomal imprinted genes we identified are not among the previously annotated imprinted genes, while 66 are known. Interestingly, 79% of the unannotated imprinted genes were found in the brain only (Figure 1E). To determine whether these tissue differences are specific to imprinting, we analyzed allele expression effects in our hybrid data that arise due to genetic differences between Cast and B6 alleles (strain effects). We statistically detected strain effects with a generalized linear model (glm) that tests for a main effect of the strain of the allele, rather than the parental origin. This approach revealed that the majority of autosomal and X-linked genes exhibit a significant bias to express either the Cast or B6 allele in each tissue (Figure 1G, 1% FDR). Thus, the tissue differences for imprinting, which involve enrichments in the brain and a paucity of effects in the liver, do not occur for strain-related genetic allelic effects.

As detailed in the Supplemental Information, we can gain insights into the sensitivity of our methods by taking advantage of the known Xm expression bias in somatic tissues of female mice (Calaway et al., 2013; Chadwick and Willard, 2005; Fowlis et al., 1991; Gregg et al., 2010a; Wang et al., 2010). Between 80%–90% of X-linked genes exhibit a maternal allele expression bias in each of the four tissue types (Figures S2A and S2B). Thus, by evaluating the proportion of maternally biased X-linked genes that are detected at the 1% FDR, we can gain insights into the sensitivity of our methods (Figure S2). For the ARN, we found that 170 (35%) of the 492 total expressed and maternally biased X-linked genes are statistically detected as imprinted (Figure S2C). In the DRN, 38 (8%) of the 499 maternally biased X-linked genes are detected. For the liver and muscle, only three (0.7%) and 12 (2%) of the 421 and 414 maternally biased X-linked genes are detected, respectively. By relaxing the cutoff to a 20% FDR, we statistically detect the maternal bias for over 70% of maternally biased X-linked genes in the ARN and DRN. Thus, at the 1% FDR cutoff, our screen is not saturated and is powered to discover imprinting effects that are similar to the most robust maternally biased X-linked genes in the ARN. The results of our transcriptome-wide imprinting analysis are presented in Table S2.

Comparison of Canonical and Noncanonical Imprinting Effects

Next, we set out to compare the prevalence of autosomal canonical versus noncanonical imprinting effects. We define canonical imprinted genes as those that have at least 99% of expression or more arising from one parental allele in at least one tissue type, indicating allele silencing (Table S3). The Illumina sequencing error rate is estimated to be ~0.01%–0.1% (Loman et al., 2012; Meacham et al., 2011), and there is a one in four chance that an error will result in a B6 read being assigned as a Cast read (or visa versa) at a given SNP site. Thus, our 1% expression cutoff for allele-silencing effects is slightly higher than the expected background of ~0.025%. We classify all imprinted genes with greater than 1% of expression arising from the repressed allele as noncanonical, since they exhibit an allele expression bias. For example, *Peg3* is a canonical imprinted gene that expresses the paternal allele and silences the maternal allele in all tissues types (Figure 2A). In contrast, *Ago2* is a noncanonical imprinted gene that exhibits a *bias* to express the maternal allele in the ARN and DRN but not the liver or muscle (Figure 2B). We found a total of 24 canonical imprinted genes that exhibit allele silencing in at least one tissue (Figure 2C). In contrast, we found 186 autosomal genes that exhibit a significant bias to express either the maternal or paternal allele, and 142 of these have not previously been annotated (Figure 2D). Therefore, noncanonical imprinting effects are ~8-fold more prevalent than strict canonical imprinting effects.

Interestingly, 79% of canonical imprinted genes are expressed and imprinted in both neural and non-neural tissues (Figure 2C), but only 12% of noncanonical imprinted genes meet these criteria (Figure 2D). We further found that 64% of noncanonical imprinted genes are specific to the brain, 20% are specific to the muscle, and only 2% are specific to the liver (Figure 2D). Particularly striking is that 37% of noncanonical imprinted genes are specific to the ARN. Thus, unlike most canonical imprinting effects, noncanonical imprinting effects are highly tissue specific.

Canonical imprinted genes are typically located in gene clusters in the genome, which are defined by shared regulatory elements. We prospectively defined "clustered" and "remote" imprinted genes according to whether they are located within 1 Mb of another imprinted gene in the genome (Tables S3 and S4). As expected, we found that 92% (22 out of 24) of canonical imprinted genes are located in a cluster (Figure 2E). In contrast, only 57% of noncanonical imprinted genes are located in a cluster, while 43% reside in remote regions of the genome that are not close to other imprinted genes (Figure 2E). In total, we found evidence for 24 candidate imprinted gene clusters on 12 chromosomes (Table S4). Our results reveal that noncanonical imprinting arises both near canonical imprinted gene clusters and in novel genomic regions.

Noncanonical Imprinting Effects Can Arise Independently from Canonical Imprinting

Our results above indicate that noncanonical imprinted genes are not simply bystanders in close proximity to canonical imprinted genes, since many reside in novel regions of the genome. Here, we further investigated the relationship between canonical and noncanonical imprinted genes. For example, *Plagl1* is a canonical PEG (paternally expressed gene) in a micro-imprinted domain that is not thought to involve a gene cluster (Iglesias-Platas et al., 2013). A neighboring gene, called *Phactr2*, has been

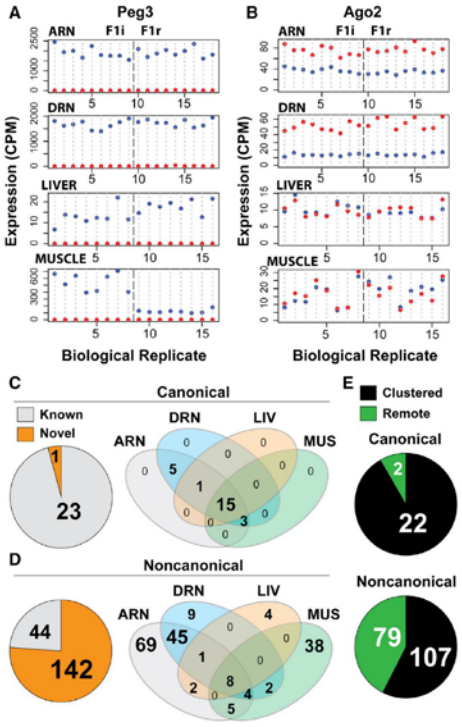


Figure 2. Comparison of Canonical and Noncanonical Imprinting Effects

(A and B) Examples of canonical (A, *Peg3*) and noncanonical (B, *Ago2*) imprinting in the ARN, DRN, liver, and muscle detected by RNA-seq. *Peg3* exhibits silencing of the maternal allele (red dots) and expression of the paternal allele (blue dots) in all tissues and biological replicates for F1i and F1r hybrid offspring (ARN, n = 9; DRN, n = 9; liver, n = 8; muscle, n = 8). *Ago2* exhibits a maternal bias in the ARN and DRN and no effect in the liver or muscle.

(C and D) Total number of known and novel canonical (C) and noncanonical (D) imprinted genes discovered and a Venn analysis of the tissues in which the genes were found.

(E) Number of canonical and noncanonical imprinted genes in clusters (clustered) compared to novel genomic regions (remote).

previously shown to exhibit noncanonical imprinting in the mouse placenta (Wang et al., 2011). In our analysis, we uncovered a maternal bias for four genes near *Plagl1*, which include *Sf3b5*, *Ltv1*, *Phactr2*, and *Fuca2* (Figure S3A). Interestingly, *Plagl1* exhibits canonical imprinting in all four tissues (Figure S3B); however, the neighboring noncanonical imprinting effects are highly tissue specific. *Sf3b5* and *Ltv1* exhibit a maternal allele bias specifically in the ARN (Figures S3C and S3D). *Phactr2* exhibits a maternal bias in the ARN and DRN, but not the liver or muscle (Figure S3E). Finally, *Fuca2* exhibits a maternal bias in the ARN and muscle, but not the DRN or liver (Figure S3F). The strength of the maternal bias for these effects does not simply decrease as a function of the distance from *Plagl1*, since *Phactr2* exhibits a stronger bias than either *Sf3b5* or *Ltv1*. Pyrosequencing confirmed that *Fuca2* exhibits a significant maternal bias in the ARN, but not the liver, in Cast × B6 and PWD/J × A/J hybrid offspring (Figure S3G). We refer readers to a second representative example at the *Inpp5f* locus (Figures S3H–S3N and Supplemental Results). A complete annotation of noncanonical imprinting effects near canonical imprinted genes is presented in Table S4. Out of the 18 gene clusters that contained canonical and noncanonical imprinted genes, 15 clusters contain maternally biased noncanonical imprinted genes only and three contain paternally biased genes only (Table S4). The *Peg3-Usp29* gene cluster has both maternally and paternally biased noncanonical imprinting effects depending on the tissue, and we validated these effects for *Clcn4*, which is maternally biased in brain and paternal in liver (Table S5).

We further identified 79 noncanonical imprinted genes in regions of the genome that do not contain other imprinted genes (Figure 2E). For example, *Nhlrc1* is located on chromosome 13 near a differentially methylated region (Xie et al., 2012). We found that *Nhlrc1* exhibits noncanonical imprinting involving a paternal bias in the ARN and DRN, but not the liver or muscle (Figures S4A and S4B). The genes surrounding *Nhlrc1* do not exhibit imprinting in any tissue (Figure S4A). Similarly, *Acrbp* exhibits a paternal bias in the ARN and DRN, but not the liver or muscle (Figures S4C and S4D). We also found similar effects in the muscle. For example, *Gbp7* (Figures S4E and S4F, chromosome 3) and *643054M08Rik* (Figures S4G and S4H, chromosome 8) exhibit a paternal and maternal bias, respectively, in muscle only. The neighboring genes do not exhibit imprinting in any of the tissues (Figures S4E and S4G). We confirmed tissue-specific imprinting for these examples and others by pyrosequencing in Cast × B6 and/or PWD/J × A/J hybrid mice (Table S5). Therefore, noncanonical imprinting effects arise independently from canonical imprinting in a highly tissue-specific manner.

Noncanonical Imprinting Effects Are Reproducible in Multiple Genetic Backgrounds

We performed pyrosequencing validations for 64 imprinted genes identified in our RNA-seq study, including 62 noncanonical imprinted genes selected from a wide range of p value cutoffs in the data. We assayed these genes in one or more tissue types, carrying out a total of 136 validation experiments involving four to eight biological replicates each. We successfully validated imprinting for 89% (57/64) of the genes tested in at least one tissue type. Out of the 136 validation experiments performed, 106 were carried out for 57 genes using Cast × B6 hybrid mice (Figure 3A; Table S5). To ascertain whether noncanonical imprinting effects are conserved across genetic backgrounds, we performed 30 further validation experiments for 23 genes with PWD/J × A/J hybrid mice (Figure 3A; Table S5).

In our validation studies using Cast × B6 hybrid mice, 17 out of 106 pyrosequencing results disagree with the RNA-seq results.

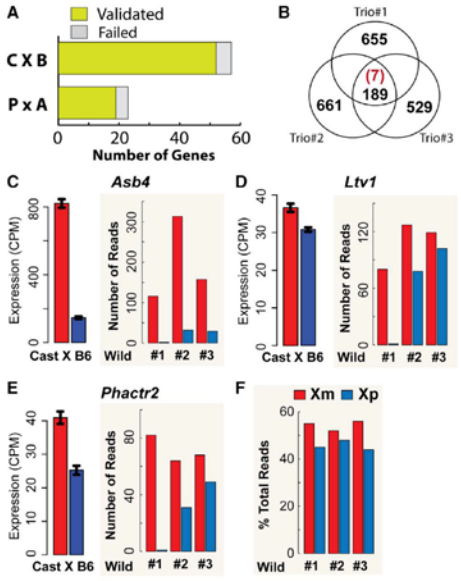


Figure 3. Noncanonical Imprinting Effects Are Highly Reproducible and Conserved in Outbred, Wild-Derived Mice

(A) Summary of pyrosequencing validation experiments in Cast × B6 (C × B) and PWD/J × A/J (P × A) hybrid offspring reveals high validation rates.

(B) Venn diagram of the number of expressed genes with SNPs for each wild-derived daughter. Out of the 189 SNP-containing genes shared between the trios, seven are imprinted genes identified in the ARN in hybrid mice.

(C–E) *Asb4*, *Ltv1*, and *Phactr2* are noncanonical MEGs with biased expression of the maternal allele (red) in Cast × B6 hybrid mice as revealed by RNA-seq. A similar maternal bias is present in each of the wild-derived daughters (1, 2, and 3).

(F) The percentage of total reads derived from Xm versus Xp alleles in the wild-derived daughters reveals an Xm expression bias.

Eight of these cases involve false-negatives in which the imprinting effect is not statistically significant by RNA-seq but is significant by pyrosequencing. Only nine cases involve potential false-positives in which imprinting is detected by RNA-seq but is not statistically significant by pyrosequencing. Finally, out of the 30 validation experiments performed in PWD/J × A/J mice, 87% (26/30 genes) agree with the RNA-seq data from Cast × B6 hybrid mice. In total, we confirmed the imprinting status for 46/50 genes tested in the ARN, 14/14 genes in the DRN, 21/25 genes in muscle, and 20/21 genes in the liver. We validated tissue-specific imprinting for 15 noncanonical imprinted genes that exhibit imprinting in the ARN, but not the liver, as well as seven genes that exhibit imprinting in the muscle only. Thus, our RNA-seq results are highly reproducible across different genetic backgrounds, and we confirmed the tissue-specific nature of noncanonical imprinting.

It is unknown whether noncanonical imprinting effects exist in wild-derived, outbred populations. To address this issue, we obtained Idaho wild-derived mice that have been maintained in captivity as an outbred colony (Miller et al., 2002). We generated three separate parent-offspring trios and harvested RNA from the hypothalamus of each parent and one daughter for each trio for analysis by RNA-seq (see the Supplemental Information). We found 189 genes that had distinguishing SNPs in all three trios and could therefore be assessed for reproducible allele-specific expression effects (Figure 3B). Out of these 189 genes, seven were identified as noncanonical imprinted genes in the ARN of F1 hybrid mice: *Asb4*, *Trappc9*, *Herc3*, *Ltv1*, *Phactr2*, *Cobl*, and *Igf2r*. In the wild-derived daughters, we found a similar noncanonical maternal bias for all of these genes except *Igf2r*, which did not exhibit imprinting (Table S6). For example, in daughter 1, *Asb4*, *Ltv1*, and *Phactr2* are almost exclusively expressed from the maternal allele, and a maternal bias is present in daughters 2 and 3 (Figures 3E and 3F). Finally, in the hybrid mice we also observed a bias to express alleles on the Xm (Figure S2), and to evaluate this phenomenon in the wild-derived mice we determined the percentage of overall expression that arises from the Xm versus the Xp in each daughter (Table S6). In all three wild-derived daughters, we found an Xm bias (Figure 5H), and the bias persists if we relax the quality score cutoff for the SNP calls to increase the total number of SNP sites examined or increase the stringency by only analyzing sites that are homozygous between the parents (data not shown). Overall, our results reveal that noncanonical imprinting effects are present in natural, outbred populations.

Tissue and Gene-Specific Imprinting Effects Arise on the X Chromosome

In females, imprinting effects can arise on the X chromosome. As noted above, at a 1% FDR, we detected imprinting effects for 198 X-linked genes, and 86% of these genes (170 genes) were detected in the ARN, compared to only 20% in the DRN, 6% in the muscle, and 1.5% in the liver (Figures 1C and 1F). Scatterplots of the mean allele bias versus the p value for imprinting for X-linked genes reveal that most X-linked genes exhibit a mean maternal bias in each of the tissues; however, the robustness of the bias appears highly gene and tissue specific (Figure 4A). Gene level imprinting effects are known to occur on the X chromosome; for example, *Xlr3b*, *Xlr4b*, and *Xlr4c* are only expressed from the Xm in some tissues (Davies et al., 2005; Raefski and O'Neill, 2005). In our study, we found that *Xlr3a*, *Xlr3c*, and *Xlr3e* exhibit the strongest maternal effects in all tissues (Figure 4A), though the statistical scores are low due to the low expression level of these genes. We further found preferential expression of the paternal *Xist* allele in all four tissues (Figure 4A), which is consistent with a bias to silence the Xp and express the Xm.

Our scatterplots also indicate X-linked genes that exhibit maternal allele expression biases in each tissue, as well as genes that do not (Figure 4A). For example, *Hmgb3* exhibits a very modest maternal bias in the ARN and DRN, and no effect in the liver and muscle (Figure 4A). In contrast, *Il13ra1* exhibits a relatively robust maternal bias in the ARN and DRN compared

OPEN ACCESS
CellPress

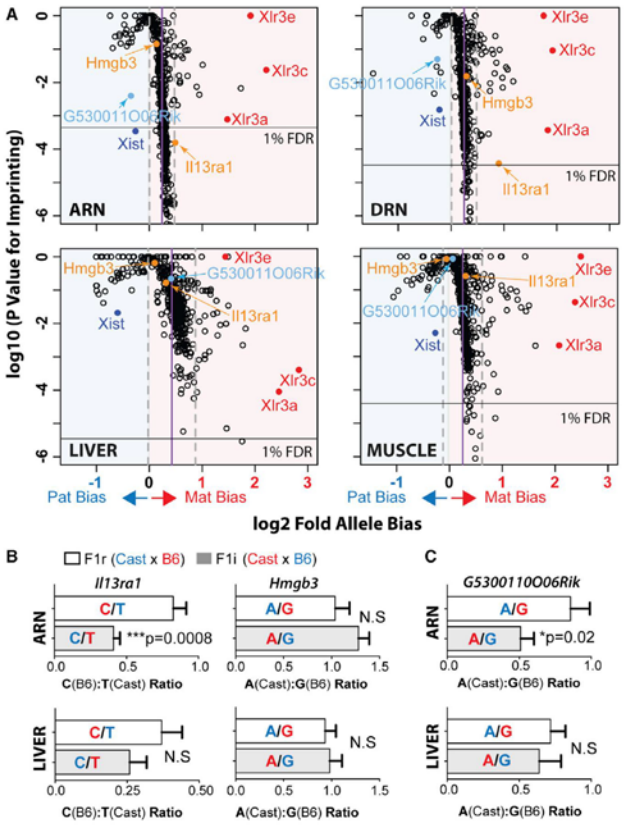


Figure 4. Tissue-Specific Imprinting Effects on the X Chromosome

(A) Scatterplots of paternal (blue side) and maternal (red side) allele expression biases (log2-fold bias) versus the p value (log10) for imprinting effects for all X-linked genes. Most X-linked genes exhibit a maternal allele expression bias (mean allele bias is indicated by the purple line; gray dashed lines indicate 1 SD from this mean). The maternally biased *Xlr* genes are indicated in red, and *Xist* is indicated in dark blue. Examples of tissue-specific X-linked imprinting effects are indicated for the MEGs, *Hmgb3* and *Il13ra1* (orange), and the PEG, *G530011O06Rik* (light blue). (B and C) Pyrosequencing validations in Cast × B6 F1 hybrid offspring. (B) *Il13ra1* and *Hmgb3* demonstrate a significant maternal bias for *Il13ra1* in the ARN, but not the liver, and *Hmgb3* does not exhibit a significant maternal bias in either tissue. (C) *G530011O06Rik* exhibits a paternal bias in the ARN, but not the liver ($n = 8$, mean ± SEM, one-tailed t test).

to *Hmgb3* (Figure 4A). Additionally, *G530011O06Rik* exhibits a paternal bias in the ARN and DRN, but not in the liver or muscle. Pyrosequencing confirmed the gene and tissue-specific noncanonical imprinting effects for these genes (Figures 4B and 4C). Pyrosequencing in Cast × B6 hybrid mice further validated brain-specific imprinting effects for the X-linked genes *Maoa*, *Bcor*, *C77370*, and *Gspt2* (Table S5). We also validated *Bcor* and *Maoa* in PWD × A/J hybrid offspring, revealing that these effects are present in different genetic backgrounds (Table S5).

Twelve genes are known to escape X-inactivation in the mouse (Yang et al., 2010), and we found that these genes also appear to exhibit tissue-specific imprinting effects (Figure S5A). For example, *Kdm6a* exhibits a modest maternal bias in the ARN, but no effect in the DRN, liver, or muscle. Pyrosequencing confirmed the maternal bias in the ARN and the absence of this bias in the DRN in Cast × B6 hybrids (Figure S5B). In PWD × A/J hybrids, we found a significant maternal bias for *Kdm6a* in both brain regions (Figure S5C). Thus, genes that are known to escape X-inactivation can exhibit a maternal allele bias, though it is unclear whether the observed effects are related to gene-specific imprinting or changes to X-inactivation escape.

Our findings detail maternal allele biases for many X-linked genes in females (Figure S2). We also tallied the total number of maternally versus paternally biased autosomal genes in each tissue (Figure S5D). Interestingly, we uncovered 107 more PEGs than MEGs on chromosome 1 in the DRN, which is a statistically significant overall paternal bias (Figure S5D, $p = 0.002$, Fisher's exact test). These results suggest biased maternal and paternal influences over X-linked and autosomal gene expression, respectively.

Noncanonical Imprinting Is Associated with Allele-Specific Histone Modifications

To ascertain whether noncanonical imprinting effects detected at the transcriptome level are associated with allele-specific chromatin modifications, we isolated chromatin from the hypothalamus of F1i and F1r Cast × B6 hybrid offspring and performed chromatin immunoprecipitation (ChIP). We targeted the transcriptionally permissive and repressive histone modifications H3K9ac and H3K9me3, respectively (Dindot et al., 2009; Singh et al., 2010), and focused on promoter regions by identifying SNPs within ±300 bp from the transcriptional start site for four canonical imprinted genes, six noncanonical imprinted genes, and one non-imprinted control gene (Table S7).

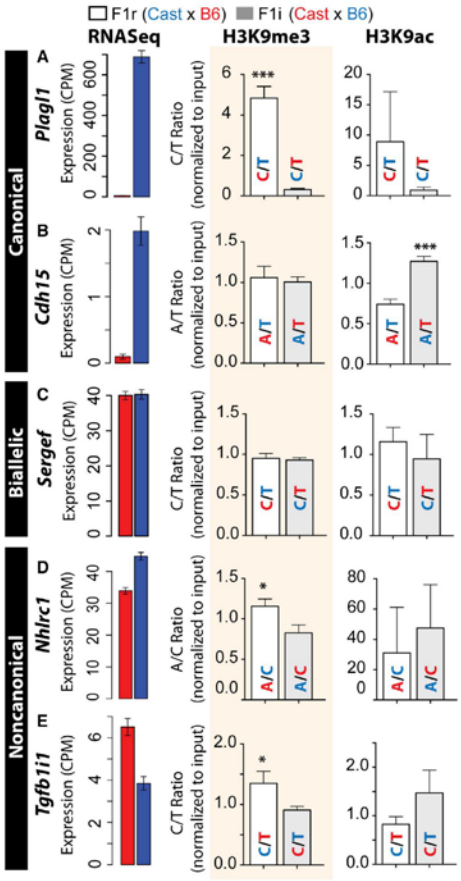


Figure 5. Chromatin Immunoprecipitation and Pyrosequencing Reveals Allele-Specific Repressive Chromatin at Noncanonical Imprinted Loci

(A and B) RNA-seq indicates that *Plagl1* (A) and *Cdh15* (B) are canonical PEGs in the ARN. ChIP-pryo analysis in Cast × B6 hybrid mice targeting SNPs sites in the promoter region for *Plagl1* reveals a significant enrichment for H3K9me3 on the repressed maternal allele and no significant enrichment for H3K9ac on either allele. In contrast, *Cdh15* exhibits a significant enrichment for H3K9ac on the expressed paternal allele and no significant enrichment for H3K9me3 on the repressed maternal allele. Enrichments are normalized to input controls.

(C) *Sergef* is a negative control gene that does not exhibit imprinting and does not exhibit maternal or paternal allele-specific enrichments for H3K9me3 or H3K9ac.

(D) *Nhlrc1* is a remote noncanonical PEG and ChIP-pyro reveals enriched H3K9me3 on the partially repressed maternal allele and no enrichment for H3K9ac on either allele.

(E) *Tgfb1i1* is a clustered noncanonical MEG and ChIP-pyro reveals a significant enrichment for H3K9me3 on the repressed paternal allele and no significant enrichment for H3K9ac on either allele. n = 6–8, two-tailed t test, mean ± SEM, ***p < 0.001, *p < 0.05.

For the canonical imprinted genes, *Plagl1* (Figure 5A), *Magel2,* and *Meg3* (Table S7), pyrosequencing revealed a significant enrichment for H3K9me3 on the silenced allele, but no enrichment for H3K9ac on the expressed allele (Figure 5A; Table S7). In contrast, for *Cdh15*, we found a significant enrichment for H3K9ac on the expressed allele but did not detect H3K9me3 enrichment on the silent allele (Figure 5B). As a negative control, we analyzed *Sergef*, which expresses the maternal and paternal alleles equally, and no allelic differences in H3K9me3 or H3K9ac enrichment were observed for this gene (Figure 5C).

Next, we analyzed six noncanonical imprinted genes, including *Nhlrc1* (PEG), *Tgfb1i1* (MEG, maternally expressed gene), *Slc25a29* (PEG), *Eif2c2* (MEG), *Trappc9* (MEG), and *Bcl2l1* (PEG). We found a significant enrichment for H3K9me3 on the repressed allele for five out of six genes (Table S7). For example, *Nhlrc1* and *Tgfb1i1* exhibit preferential expression of the paternal and maternal alleles, respectively (Figures 5D and 5E). We found a significant enrichment for H3K9me3 on the maternal allele for *Nhlrc1* (Figure 5D) and on the paternal allele for *Tgfb1i1* (Figure 5E), consistent with the repression of these alleles. Similar effects were detected for *Ago2* and *Bcl2l1*, but not for *Slc25a29* (Table S7). For *Trappc9*, we found significant H3K9ac enrichment on the maternal allele and H3K9me3 on the paternal allele (Table S7). Therefore, like canonical imprinting, noncanonical imprinting effects are associated with allele-specific chromatin modifications.

Noncanonical Imprinted Genes Exhibit Allele-Specific Expression in Subpopulations of Neurons

We tested several models to gain insights into the mechanisms underlying noncanonical imprinting effects at the cellular level. First, maternal or paternal allele-specific expression biases could be due to distinct, but overlapping transcripts from maternal versus paternal alleles (Figure 6A). However, as detailed in the Supplemental Results (and Figures S6A–S6D), we devised an approach to analyze imprinting at the transcript level and determined that *H13*, *Commd1*, *Trappc9*, *Herc3*, *Inpp5f*, *Blcap*, *Mest*, *Ube3a*, and *Gnas* are the only genes with overlapping, allele-specific isoforms (BH adjusted p value <0.01). Therefore, most noncanonical imprinting effects are not due to this phenomenon.

Alternatively, noncanonical imprinting effects could be due to (1) an allele expression bias in each cell (Figure 6B), (2) skewed random monoallelic expression effects (Figure 6C), or (3) allele silencing in a subpopulation of cells (Figure 6D). To test these possibilities, we devised an approach to resolve allele-specific expression at the cellular level using RNAscope in situ hybridization probes. Probes were designed against intronic regions to detect nascent RNA arising from each allele in the nucleus of cells in tissue cryosections of the ARN and DRN. We first evaluated this approach for the canonical MEG, *Meg3*. Our analysis is performed in isogenic female B6 mice, and we observe a single focal site of allele transcription in over 80% of positive cells,

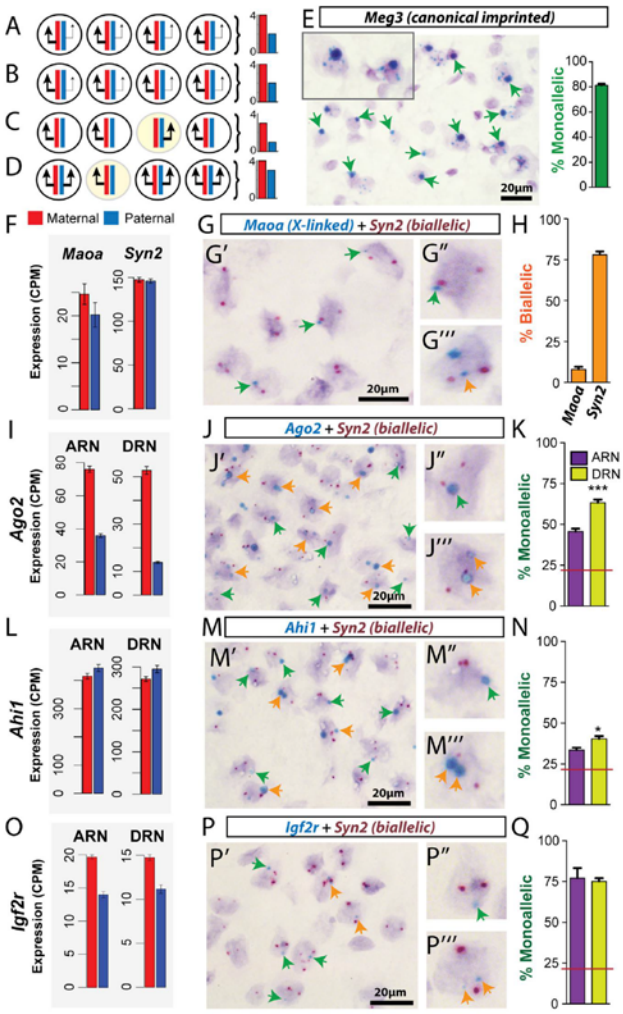


Figure 6. RNAscope Detection of Nascent RNA Reveals Subpopulations of Monoallelic Cells for Noncanonical Imprinted Genes In Vivo

(A–D) Models of noncanonical imprinting effects at the cellular level (see main text).

(E) Intronic probe targeting nascent RNA for *Meg3* (canonical imprinted) reveals monoallelic expression in cells of the DRN (green arrows). Hematoxylin nuclear counterstaining reveals nuclei in cryosections. The proportion of monoallelic cells is indicated in the bar graph (n = 3).

(F–H) RNA-seq reveals maternal bias for the X-linked gene, *Maoa*, and equal expression of the maternal and paternal alleles for *Syn2* (autosomal, ARN data shown) (F). Nascent RNA in situ (blue) in the ARN reveals largely monoallelic (P′ and P″, green arrows) and very few biallelic (P‴, orange arrows) *Maoa*$^+$ cells. In contrast, *Syn2* (dark red) expresses both alleles in most positive cells. Bar plot indicates the percentage of *Maoa*$^+$ and *Syn2*$^+$ cells that are biallelic (Q, n = 3).

(I–Q) RNA-seq data indicate maternal and paternal allele expression for *Ago2* (I), *Ahi1* (L), and *Igf2r* (O). Intronic probes reveal subpopulations of biallelic (orange arrows) and monoallelic (green arrows) cells in the ARN for *Ago2* (J), *Ahi1* (M), and *Igf2r* (P) indicated by blue staining. *Syn2* staining (biallelic control) is dark red. Bar plots indicate the percentage of monoallelic cells out of the total positive cells in the ARN and DRN (K, N, Q; n = 3–6, two-tailed t test, red line indicates *Syn2* estimated false monoallelic background). Mean ± SEM, *p < 0.05, ***p < 0.001.

consistent with canonical imprinting for this gene (Figure 6E). We further noted small speckles outside of the primary nuclear site of allele expression and demonstrate that this staining is due to some intron retention in the nascent RNA (Figures S7A–S7E). To resolve this issue, we developed methods in which probes are designed to specifically target rapidly processed introns (Figures S7A–S7E). With our approach, we can resolve allelic expression at the cellular level for any gene.

We analyzed allelic expression effects for a nonimprinted, biallelic control gene, *synapsin II* (*Syn2*), which is ubiquitously expressed in neurons and found that *Syn2* exhibits biallelic expression in ~77% of neurons (n = 3, see *Syn2* staining in Figures 6F–6H). In contrast, the X-linked gene *Maoa* exhibits biallelic expression in fewer than 8% of *Maoa*$^+$ neurons, as expected due to the effects of random X-inactivation (Figures 6F–6H). We suspected that the apparent monoallelic *Syn2*$^+$ cells are artifacts from partial nuclei cut during cryosectioning. Indeed, confocal Z-stacks of 14-μm sections revealed that 30%–35% of DAPI$^+$ nuclei are partial (Figures S7F and S7G). Therefore, based on the *Syn2* biallelic control, we conclude that our method has a background effect in which ~23% of monoallelic cells are potentially false due to sectioning artifacts.

Next, we used our approach to analyze allelic expression effects for noncanonical imprinted genes and performed double labeling with *Syn2*, as an internal biallelic control and neuron marker. We first analyzed *Ago2*, which exhibits a 2-fold maternal

allele bias in the ARN and 3-fold maternal bias in the DRN (Figure 6I). We found that 46% and 63% of *Ago2*$^+$ cells are monoallelic in the ARN and DRN, respectively, revealing significantly more monoallelic cells in the DRN, which has stronger imprinting (Figures 6J and 6K). *Ahi1* exhibits a paternal allele bias by RNA-seq, and we found that 34% and 40% of *Ahi1*$^+$ cells are monoallelic in the ARN and DRN, respectively (Figures 6L–6N). Finally, we found that *Igf2r* is a noncanonical MEG in the brain and determined that it exhibits monoallelic expression in ~75% of positive cells in the ARN and DRN (Figures 6O–6Q). Importantly, the number of monoallelic cells detected for *Ago2*, *Ahi1*, and *Igf2r* is at approximately 2-fold higher, or more, than the background of our approach (red line in Figures 6H, 6K, and 6N), and we clearly identify monoallelic effects in neurons that are biallelic for *Syn2* (Figures 6G″, 6J″, and 6M″), indicating bona fide allele-specific expression effects. To further confirm our findings, we devised a fluorescent staining and confocal imaging strategy to detect whole nuclei and determined that bona fide monoallelic and biallelic cellular subpopulations exist for cells with fully intact nuclei for both *Ahi1* (Figure S7H) and *Ago2* (Figure S7I). Based on these findings, we conclude that autosomal noncanonical imprinted genes exhibit allele-specific expression effects in subpopulations of neurons in the brain.

Noncanonical Imprinting Effects Influence the Monoamine Pathway and Offspring Phenotypes

Gene ontology analysis and literature searches provided us with insights into the biological pathways that contain imprinted genes. In the ARN and DRN, we discovered several noncanonical imprinted genes with roles in monoamine signaling, including tyrosine hydroxylase (*Th*, MEG), *Ddc* (MEG), *Maoa* (X-linked MEG), *Tgf1b1i* (MEG), and *Ahi1* (PEG), as well as known canonical imprinted genes that influence monoamine signaling, including *RasGrf1* and the snoRNA, *HBII-52* (Figures 7A and 7B) (Doe et al., 2009; Fernández-Medarde and Santos, 2011). We used pyrosequencing to evaluate the imprinting status of the noncanonical imprinted genes in major monoaminergic nuclei, as well as *Dbh* (*dopamine beta-hydroxylase*), which regulates norepinephrine (NE) synthesis, and *Tph2* (*tryptophan hydroxylase*), which regulates serotonin synthesis. We performed our analysis in the ARN, DRN, ventral tegmental area (VTA), and the locus coeruleus (LC).

Our study revealed that *Dbh* and *Tph2* are not imprinted in the LC and DRN, respectively (Figure 7C). However, *Th* exhibits a significant maternal allele bias (Figure 7D, $p < 0.05$, main effect of cross, two-way ANOVA). The bias appears in the ARN, DRN, and LC, but not in the VTA, though this brain region difference did not result in a significant interaction effect between cross and brain region (Figure 7D). On the other hand, *Ddc* exhibits a maternal bias that is significantly different between brain regions (Figure 7E, $p < 0.0001$, interaction between cross and brain region). *Ddc* imprinting is strongest in the ARN and the LC, weaker in the DRN and is not significant in the VTA (Figure 7E, Tukey HSD post hoc test). *Tgfb1i1* interacts with the dopamine transporter (DAT) (Carneiro et al., 2002) and exhibits a significant maternal bias in each of the brain regions (Figure 7F). *Ahi1* can influence serotonin signaling (Wang et al., 2012) and exhibits a significant paternal bias in each brain region (Figure 7G). Next, we used RNAscope probes to analyze allelic expression at the cellular level for *Ddc* in B6 female mice. We found that the number of *Ddc*$^+$ monoallelic cells in the ARN is significantly greater than in the VTA (Figures 7H–7J). Therefore, more monoallelic cells are detected in the brain region with stronger imprinting.

Finally, we found subpopulations of *Th*$^+$ neurons in the brain that exhibit allele-specific expression effects (Figure 7K) and sought to test whether *Th* imprinting effects influence the impact of inherited mutations on offspring behavior. We obtained *Th* mutant mice on a B6 background and generated reciprocal *Th*$^{-/+}$ (maternal deletion) and *Th*$^{+/-}$ (paternal deletion) heterozygous offspring, as well as wild-type littermates. *Th* heterozygous mice are known to exhibit reduced catecholamine levels in the brain and significant behavioral changes (Kobayashi et al., 2000), and catecholamines influence motivated behaviors. We performed open-field testing to compare *Th*$^{-/+}$ and *Th*$^{+/-}$ offspring and observed a significant effect of the parental origin of the mutation in males and females, such that offspring with a mutated maternal allele spent more time in the center of the arena compared to offspring with a mutated paternal allele (Figure 7L). No difference was observed between the wild-type littermates (Figure 7M), and the total distance traveled was not different for any genotype (data not shown). In the sucrose intake test, a hedonic measure, we found that offspring with a mutated maternal allele (*Th*$^{-/+}$) consume significantly more sucrose solution compared to those with a mutated paternal allele (*Th*$^{+/-}$) (Figure 7N). No difference was observed between the wild-type littermates (Figure 7O). In summary, noncanonical imprinting influences genes in the monoamine pathway, in some cases the effects are brain region specific, and effects at a single locus can significantly, albeit modestly, influence the effect of inherited mutations on behavior.

DISCUSSION

Studies of imprinting have uncovered important insights into the genetic architecture of complex phenotypes, human disease, and the nature of epigenetic regulatory mechanisms (Adalsteinsson and Ferguson-Smith, 2014; Lawson et al., 2013; Peters, 2014). However, little is known about noncanonical imprinting effects that manifest as allele expression biases at the tissue level. Here, we devised a sensitive RNA-seq-based approach that accurately detects both canonical and noncanonical imprinting effects in two brain regions, skeletal muscle, and liver. Our study documents 210 autosomal imprinted genes in the adult female mouse, of which 142 are noncanonical imprinted genes. In addition, we uncovered tissue-specific imprinting effects that influence X-linked genes. Our findings demonstrate that noncanonical imprinting effects are highly enriched in the brain, are expressed independently from canonical imprinting effects, are reproducible across different genetic backgrounds of inbred mice, are conserved in wild-derived mice and involve allele-specific chromatin modifications. On the basis of these results, we conclude that noncanonical imprinting is a bona fide form of epigenetic allele regulation.

We provide evidence that autosomal noncanonical imprinted genes exhibit allele-specific expression effects in discrete subpopulations of neurons in the brain. The allele-specific

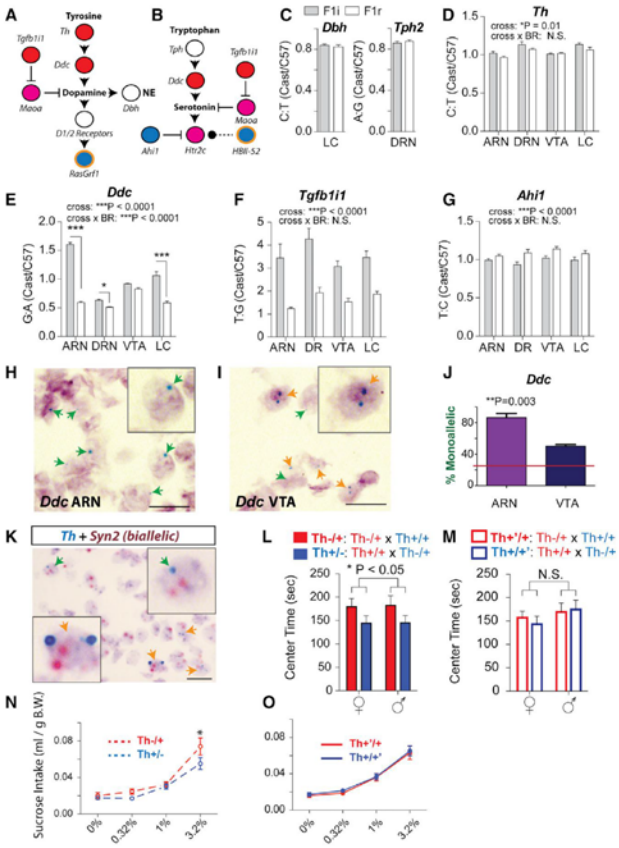


Figure 7. Noncanonical Imprinting in the Monoamine Pathway Causes Parent-of-Origin Effects for Inherited Mutations

(A and B) Summary of canonical (yellow border) and noncanonical (black border) MEGs (red) and PEGs (blue) in the catecholamine (A) and serotonin (B) pathways. X-linked MEGs indicated in pink.

(C) Pyrosequencing reveals that *Dbh* and *Tph2* are not imprinted in the LC and DRN, respectively (n = 8).

(D–G) Pyrosequencing for *Th*, *Ddc*, *Tgfb1i1*, and *Ahi1* in the ARN, DRN, VTA, and LC. All genes exhibit a significant main effect of cross revealing a noncanonical imprinting effect (n = 6–8, mean ± SEM, two-way ANOVA). Only *Ddc* exhibits a significant interaction between cross and brain region (BR), revealing brain region differences in the imprinting effect (n = 8). A Tukey HSD post-test determined that *Ddc* exhibits a significant maternal bias in the ARN, DRN, and LC, but not the VTA (*p < 0.05, **p < 0.01, **p < 0.001).

(H–J) *Ddc* nascent RNA in situ reveals monoallelic and biallelic subpopulations of neurons in the ARN (H) and VTA (I). A significantly larger proportion of *Ddc*$^{+}$ cells exhibit monoallelic expression in the ARN compared to the VTA (J, n = 3, mean ± SEM, two-tailed t test, red line indicates *Syn2* estimated false monoallelic background). Scale bar, 20 μm.

(K) *Th* nascent RNA in situ (blue) reveals monoallelic and biallelic cells in the brain (ARN shown). Biallelic *Syn2* control in dark red.

(L and M) Total time in the center region for *Th*$^{+/-}$ and $^{-/+}$ females and males in the open-field task reveals a main effect of the parental origin of the mutant allele (n = 6–8, mean ± SEM, two-way ANOVA). (M) No difference is observed between the wild-type littermates (n = 8).

(N and O) Sucrose solution intake per gram of body weight (BW) at increasing sucrose concentrations reveals a significant increase in sucrose consumption for *Th*$^{-/+}$ compared to *Th*$^{+/-}$ offspring for the 3.2% sucrose solution (n = 13, *p < 0.05, mean ± SEM, Tukey HSD post test; data for males and females are pooled since no significant sex difference was detected). (O) No difference was detected between the wild-type littermates.

expression effects could involve highly cell-type-specific canonical imprinting, or alternatively, maternally or paternally skewed random monoallelic effects (Zwemer et al., 2012) or allelic bursting (Bahar Halpern et al., 2015). The relative proportion of monoallelic cells detected in B6 mice with our approach follows the relative strength of the imprinting effect at the tissue level for *Ago2*, *Ahi1*, and *Ddc* in hybrid mice. However, for *Igf2r*, we observed more monoallelic cells than might be expected based on the strength of the imprinting effect. One explanation is that biallelic and monoallelic cells express at different levels, and therefore the tissue-level allele bias is not directly proportional to the number of monoallelic cells. Additionally, some degree of skewed random allelic expression or allelic transcriptional bursting could influence the relationship between the tissue level allele expression bias and the number of monoallelic cells. Differences between hybrid and B6 mice could also contribute.

A deeper understanding of how noncanonical imprinting impacts defined sets of neurons in the brain will be essential to determine the specific brain functions and behaviors that are influenced. We report an initial analysis of the impact of *Th* noncanonical imprinting effects on behavior in the open-field and sucrose intake tasks and identified statistically significant, but modest effects. Further insights into the identity of the neurons that exhibit allele-specific *Th* expression effects are expected to reveal the brain functions that are most strongly impacted by *Th* imprinting. Interestingly, *Th* and *Ddc* exhibit similar patterns of imprinting across the major monoaminergic nuclei of the brain, with a relatively stronger maternal allele bias in the ARN and LC, weak effects in the DRN, and no effect in the

VTA. Given that these two enzymes regulate catecholamine synthesis, we expect that their combined imprinting effects additively influence specific aspects of brain function and behavior, which is an important area for future work. Gene ontology analysis of our data also indicates other possible synergistic relationships between different canonical and noncanonical imprinted genes in the regulation of neuron differentiation, metabolic processes, and cell signaling. Interestingly, we found that over 80% of all imprinted genes are noncanonical, and therefore noncanonical imprinting is the most prevalent form of imprinting in mice. Canonical imprinting is thought to be associated with a cost due to the fact that only a single allele is expressed and therefore heterozygosity cannot buffer against deleterious mutations (Otto and Goldstein, 1992). We speculate that the evolution of highly tissue- and cell-specific noncanonical imprinting effects that function by influencing multiple genes in a pathway reduces these costs to the offspring.

Our results reveal that most canonical imprinted genes exhibit imprinting in multiple tissue types, in agreement with a recent survey (Prickett and Oakey, 2012). We provide evidence that strict canonical imprinting involving allele silencing at the tissue level, occurs for 24 genes plus nine imprinted genes that have overlapping maternal and paternal transcripts. In contrast, noncanonical imprinting effects are highly enriched in the ARN, but rare in the liver. The enrichment for noncanonical imprinting in the ARN involves both autosomal and X-linked imprinting effects. The ARN plays central roles in the regulation of hunger, metabolism, glucose homeostasis, and the neuroendocrine system (Gao and Horvath, 2007; Sternson, 2013). Several studies have defined roles for canonical imprinted genes in the hypothalamus and in the regulation of feeding, metabolism, glucose homeostasis, and the neuroendocrine axis (Ivanova and Kelsey, 2011). Thus, canonical and noncanonical imprinting effects may have evolved under similar selective pressures.

Some studies have reported loss-of-imprinting in hybrid animals (Wolf et al., 2014), raising the concern that noncanonical imprinting is due to disrupted canonical imprinting mechanisms in hybrid mice. However, there are few examples of altered imprinting in Cast × B6 hybrid mice (Korostowski et al., 2012). In our study, 32 imprinted genes annotated in public databases were not detected. Fifteen of these did not have SNPs, and most of the remaining 17 genes are known to exhibit imprinting in the placenta only. Thus, known imprinting effects are intact in our data, and we found little evidence for loss of imprinting. Most importantly, we demonstrate that noncanonical imprinting is reproducible in both Cast × B6 and PWD/J × A/J hybrids and occurs in wild-derived, outbred populations. These results provide confidence that our findings are not simply due to disrupted canonical imprinting.

This study refines and substantially extends our early observations on imprinting in the brain (Gregg et al., 2010a; 2010b). Recently, two independent groups published profiles of imprinting in mice, but they disagree in terms of their findings (Babak et al., 2015; Crowley et al., 2015). We identified significant imprinting effects for 22 of the 52 novel genes reported by Crowley et al. at a 1% FDR, and 87% of their novel genes exhibit the same direction of allele bias in our brain data, indicating strong agreement. We did not find evidence for the reported widespread paternal bias on the autosomes (Crowley et al., 2015) but uncovered a significant paternal bias on chromosome 1 in the DRN, and our internal controls indicate that other noncanonical imprinting effects likely remain to be discovered. Finally, we identified imprinting for one of the novel imprinted genes reported by Babak and colleagues (*Edn3*), and our results support their conclusion that imprinting is more prevalent in the brain compared to other tissues (Babak et al., 2015). Our study further reveals that the brain enrichment is largely driven by noncanonical imprinting effects.

EXPERIMENTAL PROCEDURES

Animals

Animal experiments were performed in accordance with protocols approved by the University of Utah Institutional Animal Care and Use Committee. C57BL/6J, CastEiJ, PWD/J, and A/J males and females were obtained from Jackson Laboratory. Idaho wild-derived mice were a gift from Dr. Steven Austad (University of Alabama). *Th* mutant mice were backcrossed onto a B6 background for at least eight generations and were a gift from Dr. Richard Palmiter (University of Washington).

RNA Isolation and RNA-Seq

The ARN, DRN, liver, and thigh muscle were microdissected from female F1 hybrid mice at 8–10 weeks of age. The ARN dissection includes the ventral medial hypothalamus. The DRN dissection includes portions of the ventral periaqueductal gray. The RNA was extracted using the RNAeasy Micro Kit (QIAGEN). RNA was pooled from four to five daughters from different litters to provide ~3 μg of total RNA for each biological replicate. Samples were prepared for RNA-seq using the TruSeq RNA sample preparation kit v2 (RS-122-2001, Illumina). Single-end 59-bp sequencing of the libraries was performed using the Illumina HiSeq 2000.

Allele-Specific ChIP

Chromatin was isolated from the hypothalamus of Cast × B6 F1 hybrid mice and chromatin immunoprecipitation for H3K9ac and H3K9me3 was performed using the Imprint Chromatin Immunoprecipitation kit according to the manufacturer's instructions (CHP1-24RXN, Sigma-Aldrich) and the following antibodies: mouse anti-H3K9ac (ab4441, Abcam) and rabbit anti-H3K9me3 (ab8898, Abcam). Results are normalized to input controls.

mRNA Preparation for Pyrosequencing

Total RNA was purified from the ARN, DRN, liver, or muscle from individual Cast × B6 F1 hybrid offspring using MicroElute Total RNA kit (R6831-02, Omega). The cDNA library was generated using the qScript cDNA supermix (P/N84034, Quanta) and oligo(dT) primers + random hexamer primers according to the manufacturer's instructions.

Pyrosequencing

Pyrosequencing to analyze allele-specific expression effects or allele-specific ChIP for specific genes was performed using the Pyromark System according to manufacturer's instructions (QIAGEN). Amplification primers and sequencing primers are provided in Tables S5 and S7. We performed four to eight biological replicates for each cross and calculated the Cast:B6 allele expression (mRNA) or enrichment (ChIP) ratio for each replicate. Using a one-tailed t test for RNA-seq validation studies and a two-tailed t test for ChIP studies, we compared the Cast:B6 allele expression or enrichment ratio between the F1i and F1r hybrid offspring to test for statistically significant differences between the two crosses.

RNAScope Allele In Situ Hybridization

RNAscope probes targeting specific introns were designed by Advanced Cell Diagnostics (http://www.acdbio.com/), and staining was performed using ACD RNAscope kits according to the manufacturer's instructions on 14-μm

tissue cryosections from female B6 mice. Probes used in this study are available for ordering from ACD.

Behavior Studies

Open-field testing was analyzed by Noldus Ethovision software (http://www.noldus.com/). Sucrose intake was determined for each concentration from 2 days of testing during the light phase with alternate cage bottle positions on each day.

ACCESSION NUMBERS

The RNA-seq data reported in this paper have been deposited to the NCBI GEO and are available under accession number GEO: GSE70484.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Results, Supplemental Experimental Procedures, seven figures, and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2015.07.017.

AUTHOR CONTRIBUTIONS

C.G. designed the study and wrote the paper. E.F. and C.G. performed the bioinformatics. W.-C.H. and T.C. performed the pyrosequencing and ChIP studies. W.-C.H. performed the wild-derived mouse experiments. C.G. and C.N.S.H. performed mouse breeding, dissections, sample prep, and data analysis. C.N.S.H., T.C., and P.J.B. performed the RNAscope allele expression studies. P.J.B. performed the behavior and monoamine pathway pyrosequencing studies.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Catherine Dulac for financial support to carry out some of the sequencing performed in this study with funds from The Klarman Foundation for Eating Disorders and Dr. Richard Palmiter for transgenic mice. We thank Dr. Ken Boucher (University of Utah) and Dr. Mark Reimers (Virginia Commonwealth University) for comments and advice on the statistical methods. We wish to thank Drs. Monica Vetter, Jan Christian, Jay Gertz, and Carl Thummel (University of Utah) for critical comments on the manuscript. C.N.S.H. is funded by a Swiss National Science Foundation Postdoctoral Fellowship. P.J.B. is funded by a postdoctoral training award from NIH training grant 5T32DK091317-02. C.G. is a New York Stem Cell Foundation Robertson-Neuroscience Investigator.

Received: December 8, 2014
Revised: June 19, 2015
Accepted: July 8, 2015
Published: July 30, 2015

REFERENCES

Adalsteinsson, B.T., and Ferguson-Smith, A.C. (2014). Epigenetic control of the genome-lessons from genomic imprinting. Genes (Basel) *5*, 635–655.

Babak, T., Deveale, B., Armour, C., Raymond, C., Cleary, M.A., van der Kooy, D., Johnson, J.M., and Lim, L.P. (2008). Global survey of genomic imprinting by transcriptome sequencing. Curr. Biol. *18*, 1735–1741.

Babak, T., DeVeale, B., Tsang, E.K., Zhou, Y., Li, X., Smith, K.S., Kukurba, K.R., Zhang, R., Li, J.B., van der Kooy, D., et al. (2015). Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. Nat. Genet. *47*, 544–549.

Bahar Halpern, K., Tanami, S., Landen, S., Chapal, M., Szlak, L., Hutzler, A., Nizhberg, A., and Itzkovitz, S. (2015). Bursty gene expression in the intact mammalian liver. Mol. Cell *58*, 147–156.

Bartolomei, M.S., and Ferguson-Smith, A.C. (2011). Mammalian genomic imprinting. Cold Spring Harb. Perspect. Biol. *3*, a002592–a002592.

Calaway, J.D., Lenarcic, A.B., Didion, J.P., Wang, J.R., Searle, J.B., McMillan, L., Valdar, W., and Pardo-Manuel de Villena, F. (2013). Genetic architecture of skewed X inactivation in the laboratory mouse. PLoS Genet. *9*, e1003853.

Carneiro, A.M., Ingram, S.L., Beaulieu, J.-M., Sweeney, A., Amara, S.G., Thomas, S.M., Caron, M.G., and Torres, G.E. (2002). The multiple LIM domain-containing adaptor protein Hic-5 synaptically colocalizes and interacts with the dopamine transporter. J. Neurosci. *22*, 7045–7054.

Chadwick, L.H., and Willard, H.F. (2005). Genetic and parent-of-origin influences on X chromosome choice in Xce heterozygous mice. Mamm. Genome *16*, 691–699.

Challis, C., Boulden, J., Veerakumar, A., Espallergues, J., Vassoler, F.M., Pierce, R.C., Beck, S.G., and Berton, O. (2013). Raphe GABAergic neurons mediate the acquisition of avoidance after social defeat. J. Neurosci. *33*, 13978–13988, 13988a.

Crowley, J.J., Zhabotynsky, V., Sun, W., Huang, S., Pakatci, I.K., Kim, Y., Wang, J.R., Morgan, A.P., Calaway, J.D., Aylor, D.L., et al. (2015). Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. Nat. Genet. *47*, 353–360.

Davies, W., Isles, A., Smith, R., Karunadasa, D., Burrmann, D., Humby, T., Ojarikre, O., Biggin, C., Skuse, D., Burgoyne, P., and Wilkinson, L. (2005). Xlr3b is a new imprinted candidate for X-linked parent-of-origin effects on cognitive function in mice. Nat. Genet. *37*, 625–629.

DeVeale, B., van der Kooy, D., and Babak, T. (2012). Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective. PLoS Genet. *8*, e1002600.

Dindot, S.V., Person, R., Strivens, M., Garcia, R., and Beaudet, A.L. (2009). Epigenetic profiling at mouse imprinted gene clusters reveals novel epigenetic and genetic features at differentially methylated regions. Genome Res. *19*, 1374–1383.

Doe, C.M., Relkovic, D., Garfield, A.S., Dalley, J.W., Theobald, D.E.H., Humby, T., Wilkinson, L.S., and Isles, A.R. (2009). Loss of the imprinted snoRNA mbii-52 leads to increased 5htr2c pre-RNA editing and altered 5HT2CR-mediated behaviour. Hum. Mol. Genet. *18*, 2140–2148.

Dölen, G., Darvishzadeh, A., Huang, K.W., and Malenka, R.C. (2013). Social reward requires coordinated activity of nucleus accumbens oxytocin and serotonin. Nature *501*, 179–184.

Fernández-Medarde, A., and Santos, E. (2011). The RasGrf family of mammalian guanine nucleotide exchange factors. Biochim. Biophys. Acta *1815*, 170–188.

Fowlis, D.J., Ansell, J.D., and Micklem, H.S. (1991). Further evidence for the importance of parental source of the Xce allele in X chromosome inactivation. Genet. Res. *58*, 63–65.

Gao, Q., and Horvath, T.L. (2007). Neurobiology of feeding and energy expenditure. Annu. Rev. Neurosci. *30*, 367–398.

Goncalves, A., Leigh-Brown, S., Thybert, D., Stefflova, K., Turro, E., Flicek, P., Brazma, A., Odom, D.T., and Marioni, J.C. (2012). Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. Genome Res. *22*, 2376–2384.

Gregg, C., Zhang, J., Butler, J.E., Haig, D., and Dulac, C. (2010a). Sex-specific parent-of-origin allelic expression in the mouse brain. Science *329*, 682–685.

Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G.P., Haig, D., and Dulac, C. (2010b). High-resolution analysis of parent-of-origin allelic expression in the mouse brain. Science *329*, 643–648.

Haig, D. (2000). The kinship theory of genomic imprinting. Annu. Rev. Ecol. Syst. *31*, 9–32.

Huguet, G., Ey, E., and Bourgeron, T. (2013). The genetic landscapes of autism spectrum disorders. Annu. Rev. Genomics Hum. Genet. *14*, 191–213.

Iglesias-Platas, I., Court, F., Camprubi, C., Sparago, A., Guillaumet-Adkins, A., Martin-Trujillo, A., Riccio, A., Moore, G.E., and Monk, D. (2013). Imprinting at the PLAGL1 domain is contained within a 70-kb CTCF/cohesin-mediated non-allelic chromatin loop. Nucleic Acids Res. *41*, 2171–2179.

Ivanova, E., and Kelsey, G. (2011). Imprinted genes and hypothalamic function. J. Mol. Endocrinol. *47*, R67–R74.

Khatib, H. (2007). Is it genomic imprinting or preferential expression? BioEssays *29*, 1022–1028.

Kobayashi, K., Noda, Y., Matsushita, N., Nishii, K., Sawada, H., Nagatsu, T., Nakahara, D., Fukabori, R., Yasoshima, Y., Yamamoto, T., et al. (2000). Modest neuropsychological deficits caused by reduced noradrenaline metabolism in mice heterozygous for a mutated tyrosine hydroxylase gene. J. Neurosci. *20*, 2418–2426.

Korostowski, L., Sedlak, N., and Engel, N. (2012). The Kcnq1ot1 long non-coding RNA affects chromatin conformation and expression of Kcnq1, but does not regulate its imprinting in the developing heart. PLoS Genet. *8*, e1002956.

Lawson, H.A., Cheverud, J.M., and Wolf, J.B. (2013). Genomic imprinting and parent-of-origin effects on complex traits. Nat. Rev. Genet. *14*, 609–617.

Loman, N.J., Misra, R.V., Dallman, T.J., Constantinidou, C., Gharbia, S.E., Wain, J., and Pallen, M.J. (2012). Performance comparison of benchtop high-throughput sequencing platforms. Nat. Biotechnol. *30*, 434–439.

Lowry, C.A., Hale, M.W., Evans, A.K., Heerkens, J., Staub, D.R., Gasser, P.J., and Shekhar, A. (2008). Serotonergic systems, anxiety, and affective disorder: focus on the dorsomedial part of the dorsal raphe nucleus. Ann. N Y Acad. Sci. *1148*, 86–94.

Meacham, F., Boffelli, D., Dhahbi, J., Martin, D.I., Singer, M., and Pachter, L. (2011). Identification and correction of systematic error in high-throughput sequence data. BMC Bioinformatics *12*, 451.

Michelsen, K.A., Schmitz, C., and Steinbusch, H.W.M. (2007). The dorsal raphe nucleus–from silver stainings to a role in depression. Brain Res. Brain Res. Rev. *55*, 329–342.

Miller, R.A., Harper, J.M., Dysko, R.C., Durkee, S.J., and Austad, S.N. (2002). Longer life spans and delayed maturation in wild-derived mice. Exp. Biol. Med. (Maywood) *227*, 500–508.

Monti, J.M. (2010). The structure of the dorsal raphe nucleus and its relevance to the regulation of sleep and wakefulness. Sleep Med. Rev. *14*, 307–317.

Otto, S.P., and Goldstein, D.B. (1992). Recombination and the evolution of diploidy. Genetics *131*, 745–751.

Peters, J. (2014). The role of genomic imprinting in biology and disease: an expanding view. Nat. Rev. Genet. *15*, 517–530.

Prickett, A.R., and Oakey, R.J. (2012). A survey of tissue-specific genomic imprinting in mammals. Mol. Genet. Genomics *287*, 621–630.

Raefski, A.S., and O'Neill, M.J. (2005). Identification of a cluster of X-linked imprinted genes in mice. Nat. Genet. *37*, 620–624.

Schulz, R., Woodfine, K., Menheniott, T.R., Bourc'his, D., Bestor, T., and Oakey, R.J. (2008). WAMIDEX: a web atlas of murine genomic imprinting and differential expression. Epigenetics *3*, 89–96.

Singh, P., Cho, J., Tsai, S.Y., Rivas, G.E., Larson, G.P., and Szabó, P.E. (2010). Coordinated allele-specific histone acetylation at the differentially methylated regions of imprinted genes. Nucleic Acids Res. *38*, 7974–7990.

Sternson, S.M. (2013). Hypothalamic survival circuits: blueprints for purposive behaviors. Neuron *77*, 810–824.

Wang, Q.P., and Nakai, Y. (1994). The dorsal raphe: an important nucleus in pain modulation. Brain Res. Bull. *34*, 575–585.

Wang, X., Sun, Q., McGrath, S.D., Mardis, E.R., Soloway, P.D., and Clark, A.G. (2008). Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. PLoS ONE *3*, e3839.

Wang, X., Soloway, P.D., and Clark, A.G. (2010). Paternally biased X inactivation in mouse neonatal brain. Genome Biol. *11*, R79.

Wang, X., Soloway, P.D., and Clark, A.G. (2011). A survey for novel imprinted genes in the mouse placenta by mRNA-seq. Genetics *189*, 109–122.

Wang, H., Huang, Z., Huang, L., Niu, S., Rao, X., Xu, J., Kong, H., Yang, J., Yang, C., Wu, D., et al. (2012). Hypothalamic Ahi1 mediates feeding behavior through interaction with 5-HT2C receptor. J. Biol. Chem. *287*, 2237–2246.

Wolf, J.B., Oakey, R.J., and Feil, R. (2014). Imprinted gene expression in hybrids: perturbed mechanisms and evolutionary implications. Heredity (Edinb) *113*, 167–175.

Xie, W., Barr, C.L., Kim, A., Yue, F., Lee, A.Y., Eubanks, J., Dempster, E.L., and Ren, B. (2012). Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. Cell *148*, 816–831.

Yang, F., Babak, T., Shendure, J., and Disteche, C.M. (2010). Global survey of escape from X inactivation by RNA-sequencing in mouse. Genome Res. *20*, 614–622.

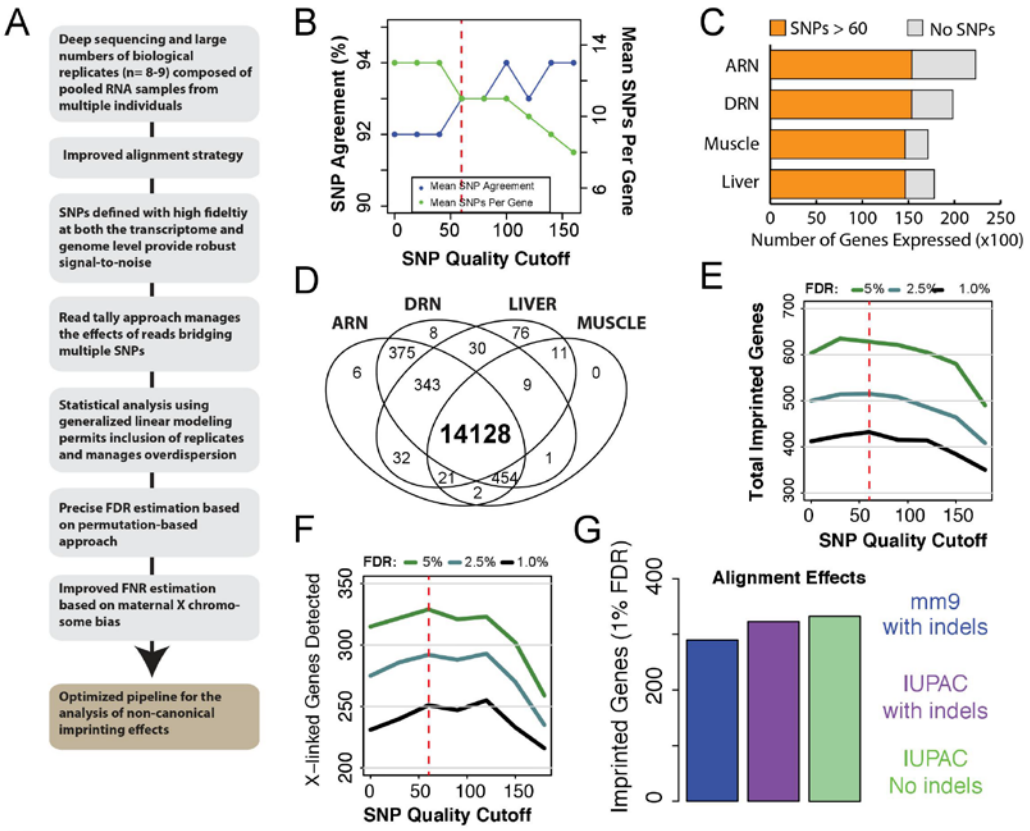
Zwemer, L.M., Zak, A., Thompson, B.R., Kirby, A., Daly, M.J., Chess, A., and Gimelbrant, A.A. (2012). Autosomal monoallelic expression in the mouse. Genome Biol. *13*, R10.

Cell Reports
Supplemental Information

Noncanonical Genomic Imprinting Effects in Offspring

Paul J. Bonthuis, Wei-Chao Huang, Cornelia N. Stacher Hörndli, Elliott Ferris, Tong Cheng, and Christopher Gregg

Figure S1



H

REAL STUDY	Cross:	F1i	F1i	F1i	F1i	F1i	F1i	F1i	F1i	F1i	F1i	F1i	F1i	F1i	F1i	F1i	F1i	F1i	F1i
	Sample:	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9
	Allele:	Cast	B6	Cast	B6	Cast	B6	Cast	B6	Cast	B6	Cast	B6	Cast	B6	Cast	B6	Cast	B6
	Parent:	Mat	Pat	Mat	Pat	Mat	Pat	Mat	Pat	Mat	Pat	Mat	Pat	Mat	Pat	Mat	Pat	Mat	Pat
	Cross:	F1r	F1r	F1r	F1r	F1r	F1r	F1r	F1r	F1r	F1r	F1r	F1r	F1r	F1r	F1r	F1r	F1r	F1r
	Sample:	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9
	Allele:	Cast	B6	Cast	B6	Cast	B6	Cast	B6	Cast	B6	Cast	B6	Cast	B6	Cast	B6	Cast	B6
	Parent:	Pat	Mat	Pat	Mat	Pat	Mat	Pat	Mat	Pat	Mat	Pat	Mat	Pat	Mat	Pat	Mat	Pat	Mat

PERMUTED	Cross:	F1i	F1i	F1r	F1r	F1i	F1i	F1r	F1r	F1i	F1i	F1r	F1r	F1i	F1i	F1r	F1r	F1i	F1i
	Sample:	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9
	Allele:	Cast	B6	Cast	B6	Cast	B6	Cast	B6	Cast	B6	Cast	B6	Cast	B6	Cast	B6	Cast	B6
	Parent:	Mat	Pat	Pat	Mat	Mat	Pat	Pat	Mat	Mat	Pat	Pat	Mat	Mat	Pat	Pat	Mat	Mat	Pat
	Cross:	F1r	F1r	F1i	F1i	F1r	F1r	F1i	F1i	F1r	F1r	F1i	F1i	F1r	F1r	F1i	F1i	F1r	F1r
	Sample:	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9
	Allele:	B6	Cast	B6	Cast	B6	Cast	B6	Cast	B6	Cast	B6	Cast	B6	Cast	B6	Cast	B6	Cast
	Parent:	Pat	Mat	Mat	Pat	Pat	Mat	Mat	Pat	Pat	Mat	Mat	Pat	Pat	Mat	Mat	Pat	Pat	Mat

Figure S2

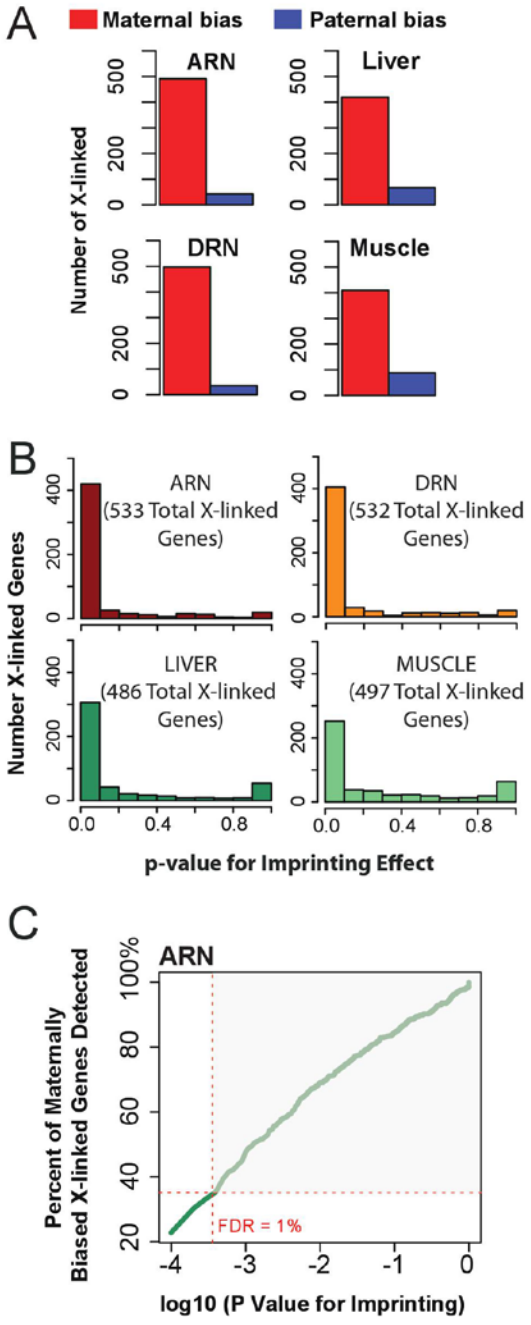

A
Maternal bias
Paternal bias
ARN
Liver
DRN
Muscle
Number of X-linked
0
200
500
B
ARN
(533 Total X-linked Genes)
DRN
(532 Total X-linked Genes)
LIVER
(486 Total X-linked Genes)
MUSCLE
(497 Total X-linked Genes)
Number X-linked Genes
0
200
400
0.0
0.4
0.8
p-value for Imprinting Effect
C
ARN
Percent of Maternally Biased X-linked Genes Detected
20
40
60
80
100%
FDR = 1%
-4
-3
-2
-1
0
log10 (P Value for Imprinting)

Figure S3

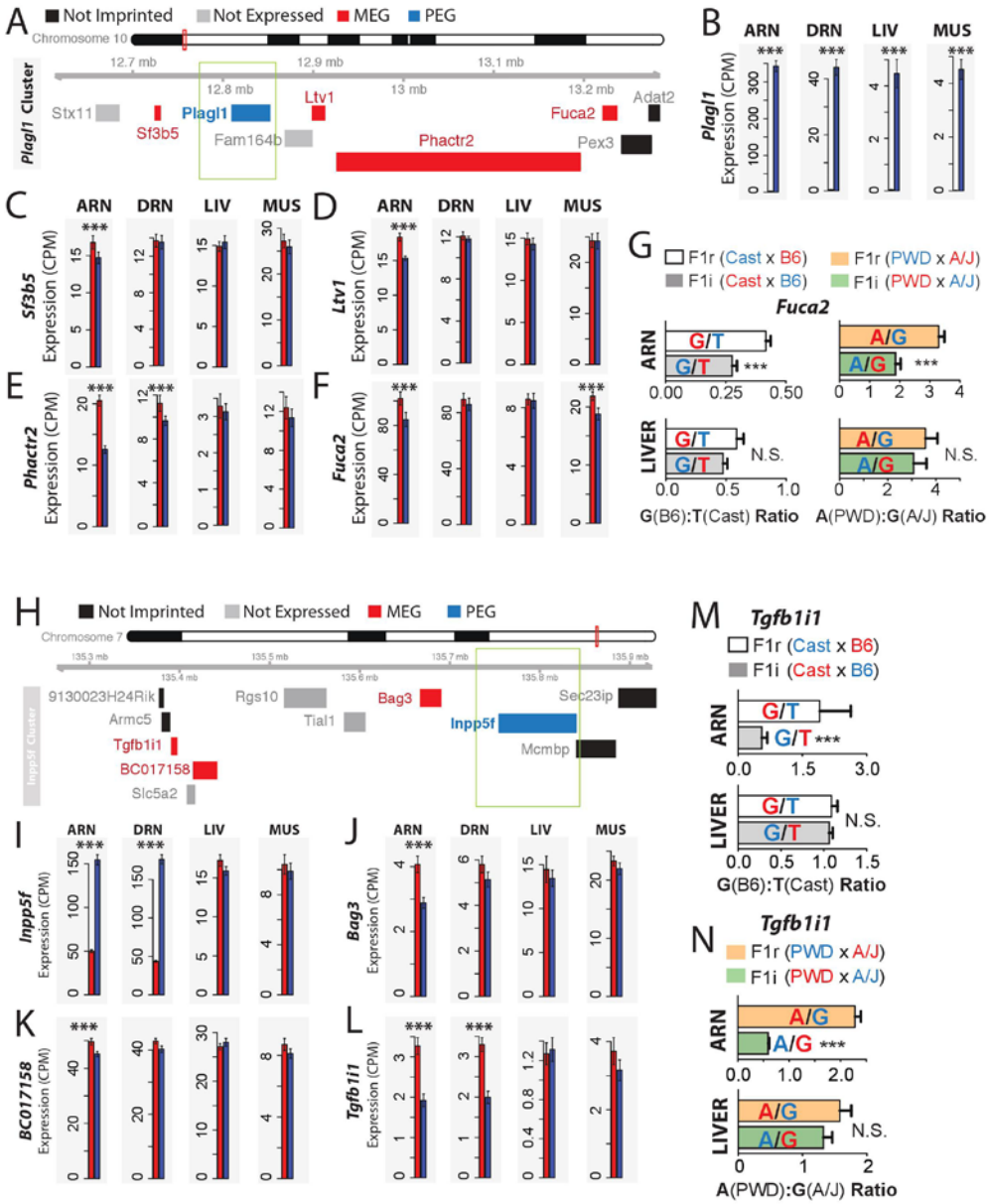

A
Not Imprinted
Not Expressed
MEG
PEG
Chromosome 10
Plagl1 Cluster
Stx11
Sf3b5
Plagl1
Ltv1
Fam164b
Phactr2
Fuca2
Pex3
Adat2
B
ARN
DRN
LIV
MUS
Plagl1
Expression (CPM)
C
Sf3b5
D
Ltv1
E
Phactr2
F
Fuca2
G
F1r (Cast x B6)
F1i (Cast x B6)
F1r (PWD x A/J)
F1i (PWD x A/J)
Fuca2
G/T
A/G
N.S.
G(B6):T(Cast) Ratio
A(PWD):G(A/J) Ratio
H
Chromosome 7
Inpp5f Cluster
9130023H24Rik
Armc5
Tgfb1i1
BC017158
Slc5a2
Rgs10
Tial1
Bag3
Inpp5f
Mcmbp
Sec23ip
I
Inpp5f
J
Bag3
K
BC017158
L
Tgfb1i1
M
Tgfb1i1
N
Tgfb1i1

Figure S4

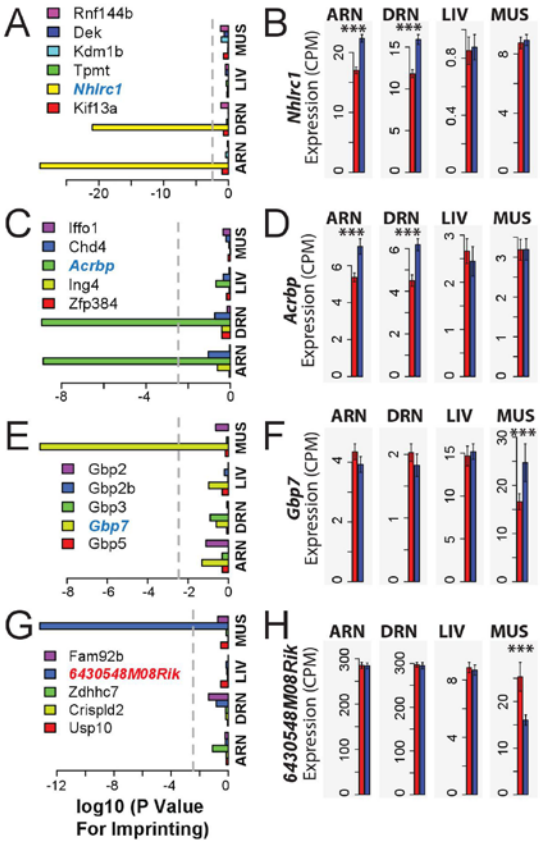

A
Rnf144b
Dek
Kdm1b
Tpmt
Nhlrc1
Kif13a
ARN DRN LIV MUS
-20 -10 0
B
ARN DRN LIV MUS
Nhlrc1
Expression (CPM)
C
Iffo1
Chd4
Acrbp
Ing4
Zfp384
-8 -4 0
D
Acrbp
E
Gbp2
Gbp2b
Gbp3
Gbp7
Gbp5
-8 -6 -4 -2 0
F
Gbp7
G
Fam92b
6430548M08Rik
Zdhhc7
Crispld2
Usp10
-12 -8 -4 0
log10 (P Value
For Imprinting)
H
6430548M08Rik
***

Figure S5

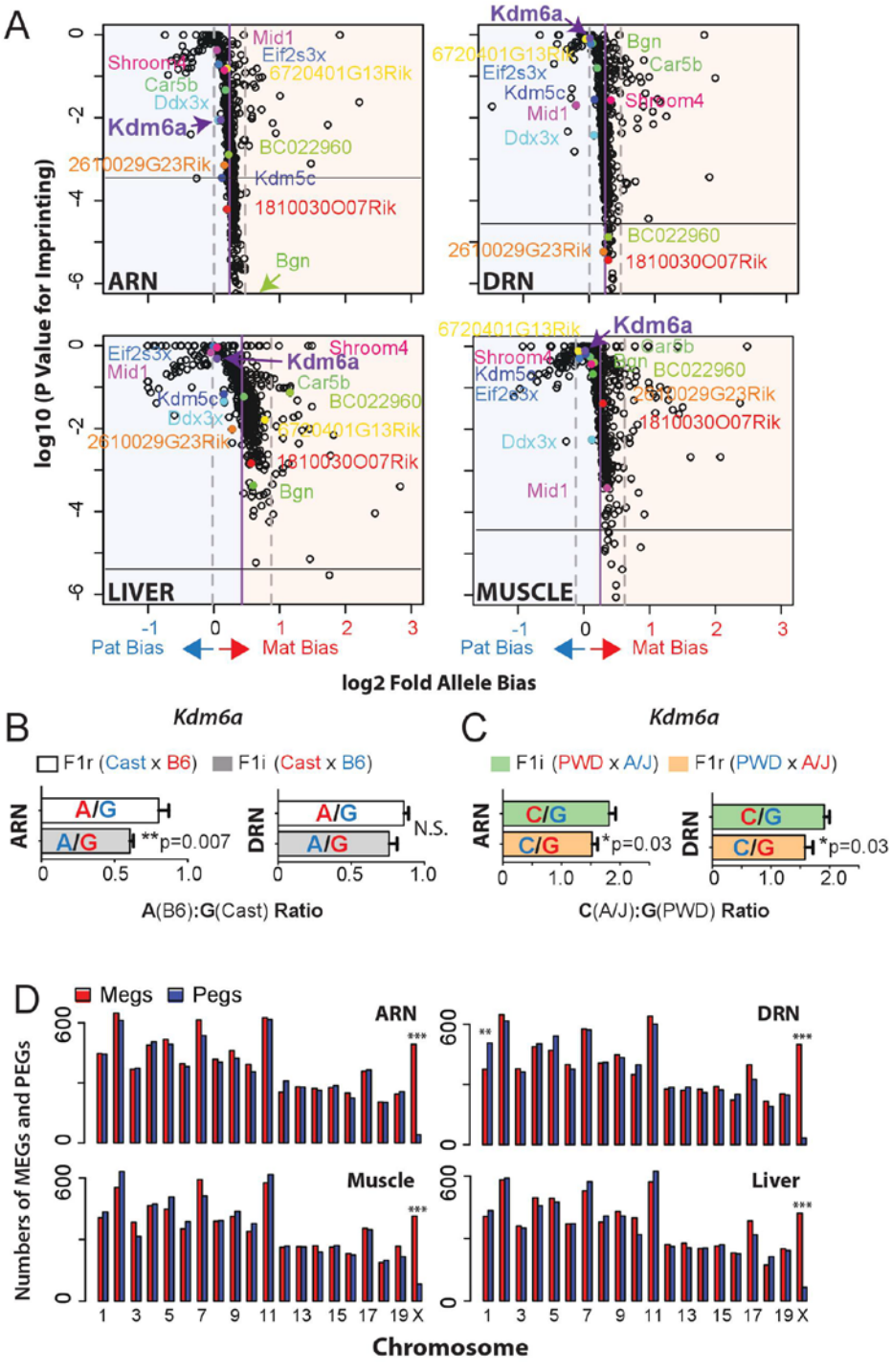

A
log10 (P Value for Imprinting)
ARN
DRN
LIVER
MUSCLE
Kdm6a
Mid1
Eif2s3x
6720401G13Rik
Shroom4
Car5b
Ddx3x
BC022960
2610029G23Rik
Kdm5c
1810030O07Rik
Bgn
Pat Bias
Mat Bias
log2 Fold Allele Bias
B
Kdm6a
F1r (Cast x B6)
F1i (Cast x B6)
A/G
**p=0.007
N.S.
A(B6):G(Cast) Ratio
C
F1i (PWD x A/J)
F1r (PWD x A/J)
C/G
*p=0.03
C(A/J):G(PWD) Ratio
D
Megs
Pegs
Numbers of MEGs and PEGs
Muscle
Liver
Chromosome

Figure S6

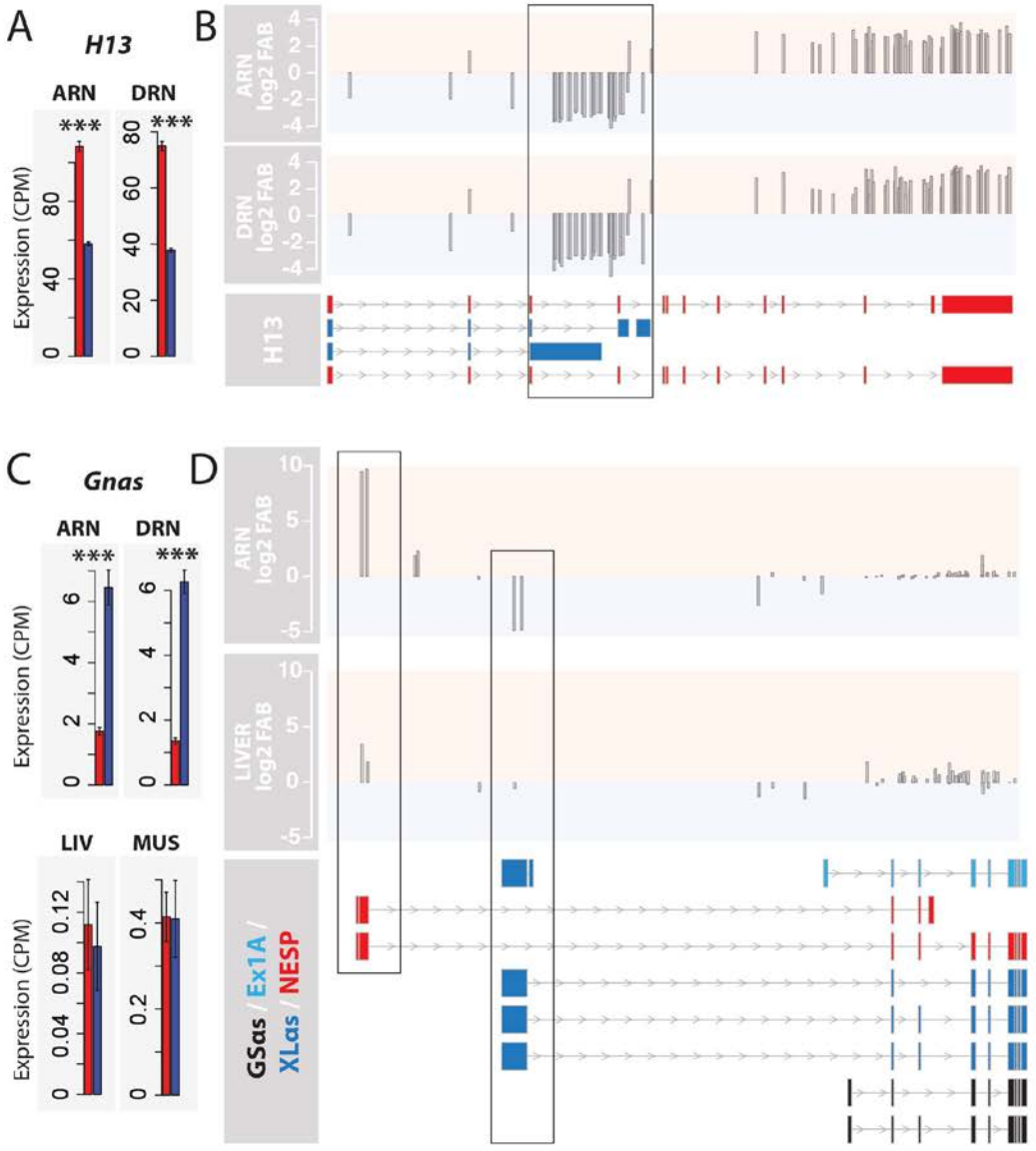
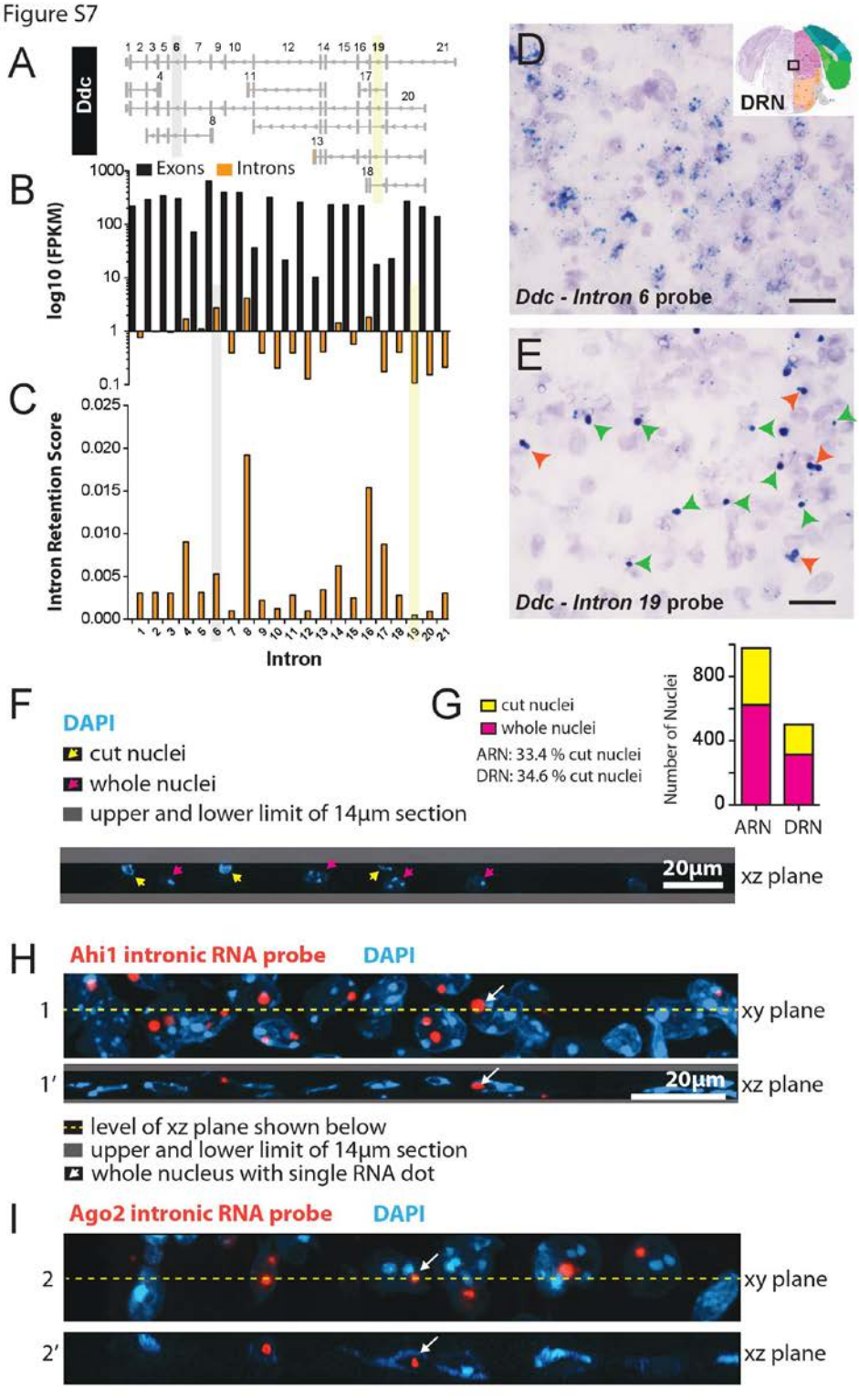

A
H13
ARN
DRN
***
***
Expression (CPM)
0
40
80
0
20
40
60
80
B
ARN log2 FAB
DRN log2 FAB
H13
4
2
0
-2
-4
C
Gnas
ARN
DRN
***
***
Expression (CPM)
0
2
4
6
LIV
MUS
0
0.04
0.08
0.12
0
0.2
0.4
D
ARN log2 FAB
LIVER log2 FAB
10
5
0
-5
GSas / Ex1A /
XLas / NESP

Figure S7


A
Ddc
B
Exons
Introns
log10 (FPKM)
C
Intron Retention Score
Intron
D
DRN
Ddc - Intron 6 probe
E
Ddc - Intron 19 probe
F
DAPI
cut nuclei
whole nuclei
upper and lower limit of 14µm section
20µm
xz plane
G
cut nuclei
whole nuclei
ARN: 33.4 % cut nuclei
DRN: 34.6 % cut nuclei
Number of Nuclei
ARN
DRN
H
Ahi1 intronic RNA probe
DAPI
xy plane
xz plane
level of xz plane shown below
upper and lower limit of 14µm section
whole nucleus with single RNA dot
I
Ago2 intronic RNA probe
DAPI
xy plane
xz plane

Supplementary Figure Legends

Figure S1 (related to Figure 1 and Experimental Procedures). An optimized analysis pipeline for the detection of noncanonical imprinting effects.

(**A**) Schematic overview of optimized analysis pipeline for the detection of noncanonical imprinting effects.

(**B**) A SNP quality score cutoff of 60 yields a high agreement between SNPs in terms of the direction of the allele bias for SNPs in the same imprinted gene (green line; 93% of SNPs indicate the same direction of allele bias). The mean number of SNPs per imprinted gene is eleven at the quality score cutoff of 60 (blue line).

(**C**) The number of genes expressed in each tissue type and the proportion with SNPs that have a quality score greater than or equal to 60.

(**D**) A Venn diagram of the number of genes assessed for imprinting in each tissue type.

(**E** and **F**) A SNP quality score cutoff of 60 maximizes the sensitivity to detect imprinting effects at the 1% FDR cutoff (E) and increases the sensitivity to detect imprinting for X-linked genes compared to less stringent SNP quality cutoffs (F).

(**G**) Total number of imprinted genes discovered at a 1% FDR following alignment to the mm9 genome (blue bar) versus alignment to an unbiased reference genome with IUPAC codes for Cast versus B6 SNP sites (purple and green bars). The effect of excluding indels and SNPs within one read length of an indel for the unbiased alignment is also shown (green) and further increases the sensitivity to detect imprinting effects in the data.

(**H**) Schematic depicting the study design (REAL) and permuted study design (PERMUTED). The permuted data intermingles the F1i and F1r samples to scramble the parental origin of the samples relative to the glm design matrix for the statistical detection of imprinting effects. Imprinting effects detected in the permuted condition are false. The data is permuted according to this pattern 500 times to determine the mean number of false positives at different p-value cutoffs.

Figure S2 (related to Figure 2). Maternally biased expression of X-linked genes in adult female mouse tissues.

(**A**) Plots of the number of X-linked genes exhibiting a mean maternal (red) versus paternal (blue) allele expression bias reveal that most X-linked genes exhibit a maternal bias in the ARN, DRN, liver and muscle.

(**B**) Plots of the p-values for imprinting effects for X-linked genes reveal that the maternal bias is statistically detected in each tissue at the gene level, as demonstrated by shifting of the p-value distribution toward low p-values. Thus, most X-linked genes exhibit a statistically detectable maternal bias.

(**C**) A plot of the percentage of all maternally biased X-linked genes expressed in the ARN that are discovered as imprinted at different p-value cutoffs. The p-value cutoff that yields a 1% FDR detects 35% of all maternally biased X-linked genes. Thus, 65% of maternally biased X-linked genes are not detected (grey box) indicating that false negatives (type II statistical errors) exist in the datasets.

Figure S3 (related to Figure 2). Noncanonical imprinting effects near the canonical imprinted genes *Plagl1* and *Inpp5f* exhibit distinct and tissue-specific imprinting effects.

(**A**) Schematic of the *Plagl1* imprinted gene cluster on chromosome 10. Green box indicates genes that exhibit canonical imprinting and noncanonical imprinted genes are indicated by orange type.

(**B-F**) Bar plots of the RNASeq results for genes in the *Plagl1* cluster indicating the mean expression levels of the maternal and paternal alleles in the four tissue types (error bars indicate standard error; asterisks indicate imprinting effect at the 1% FDR cutoff). (C) *Plagl1* exhibits canonical imprinting involving silencing of the maternal allele in all four tissues. *Sf3b5* (C) and *Ltv1* (D) exhibit a noncanonical maternal bias in the ARN only, while *Phactr2* (E) exhibits a maternal bias in the ARN and DRN. *Fuca2* (F) exhibits a maternal bias in the ARN and muscle. CPM, read counts per million.

(**G**) Pyrosequencing validation for *Fuca2* confirms a maternal bias in the ARN, but not the liver, in Cast x B6 hybrid offspring and PWD/J x A/J hybrid offspring. Pyrosequencing primers target a G/T SNP in the B6:Cast allele expression analysis (grey and white barplots) and an A/G SNP site for the PWD/J:A/J allele expression analysis (coloured barplots) (red, maternal allele; blue, paternal allele). The ratio for the expression of the G:T and A:G alleles is compared between the F1i and F1r crosses for Cast x B6 and PWD/J x A/J mice, respectively. A significant maternal bias is detected in the ARN, but not the liver in both hybrid mouse strains (N=8, mean ± SEM, one tailed t-test, ***p<0.001).

(**H**) Schematic of the *Inpp5f* imprinted gene cluster on chromosome 7. Green box indicates the known PEG, *Inpp5f,* and orange text indicates non-canonical imprinted genes.

(**I-L**) Bar plots of the RNASeq results for imprinted genes indicating the mean expression levels of the maternal and paternal alleles across all biological replicates in the four tissue types (error bars indicate standard error; asterisks indicate cases in which an imprinting effect is detected at the 1% FDR cutoff). *Inpp5f* (B) exhibits a robust paternal bias in the ARN and DRN, but not the liver (LIV) or muscle (MUS). *Bag3* (C) and *BC017158* (D) exhibit a significant maternal bias in the ARN only, while *Tgfb1i1* (E) exhibits a maternal bias in the ARN and DRN, but not the liver or muscle. CPM, read counts per million.

(**M** and **N**) Pyrosequencing validation of the maternal bias for *Tgfb1i1* in the ARN, but not the liver, in Cast x B6 hybrid offspring (F) and PWD/J x A/J hybrid offspring (G). Pyrosequencing primers targeted a G/T SNP in the *Tgfb1i1* gene for the B6:Cast allele expression analysis (grey and white barplots) and an A/G SNP site for the PWD/J:A/J allele expression analysis (coloured barplots) (red, maternal allele; blue, paternal allele). The ratio for the expression of the G:T and A:G alleles is compared between the F1i and F1r crosses for Cast x B6 and PWD/J x A/J mice, respectively. A significant maternal bias is detected in the ARN, but not the liver in both hybrid mouse strains (N=8, one tailed t-test, ***p<0.001).

Figure S4 (related to Figure 2). Tissue-specific noncanonical imprinting effects in novel genomic regions.

(**A-H**) RNASeq expression data reveals that *Nhlrc1* (A and B) and *Acrbp* (C and D) exhibit a significant paternal allele expression bias in the ARN and DRN, but not the liver or muscle. Genes neighboring *Nhlrc1* (A) or *Acrbp* (C) do not exhibit significant imprinting effects (grey dashed line indicates the p-value that yields a 5% FDR for the ARN as a reference). CPM, read counts per million. *Gbp7* (E and F) and *6430548MO8Rik* (G and H) exhibit noncanonical imprinting effects in the muscle. Neighboring genes do not exhibit significant imprinting effects.

Figure S5 (related to Figure 4). Imprinting effects for genes known to escape X-inactivation and total MEGs and PEGs on each chromosome.

(**A**) Scatterplots of paternal (blue side) and maternal (red side) allele expression biases (log2-fold bias) versus the p-value (log10) for imprinting effects for all X-linked genes in the ARN, DRN, liver and muscle. Most X-linked genes exhibit a maternal allele expression bias in each tissue, but the robustness of these effects are highly gene and tissue specific. The mean allele bias is indicated by the purple line, and grey dashed lines indicate one standard deviation from this mean. Genes known to escape X-inactivation are indicated by colored dots. Some genes that escape X-inactivation exhibit evidence for tissue-specific imprinting, such as *Kdm6a*, which exhibits a trend for imprinting in the ARN, but not the DRN, liver or muscle. Black horizontal line indicates 1% FDR cutoff. Note that *Bgn* exhibits a highly significant imprinting effect in the ARN and is off the scale of the graph.

(**B** and **C**) Pyrosequencing validations for *Kdm6a* in Cast x B6 F1 hybrid offspring reveal a significant maternal bias in the ARN, but not the DRN, consistent with the RNASeq data. Pyrosequencing in PWD x A/J hybrids reveals a significant maternal bias in both the ARN and DRN. (n=8, ***P<0.001, one tailed t-test).

(**D**) Bar plots comparing the total number of MEGs and PEGs on each chromosome independent of a statistical cutoff in the ARN, DRN, liver and muscle. A significant paternal bias is observed on

chromosome 1 in the DRN and a maternal bias is observed on the X chromosome in all tissues. **P<0.01, ***P<0.001, Fisher's Exact Test.

Figure S6 (related to Figure 6). Resolution of transcript level imprinting effects by RNASeq.

(**A** and **B**) Bar plots of the RNASeq data for *H13* in the ARN and DRN demonstrate that a maternal bias exists at the gene level (A). Plots of the log2 fold allele bias (FAB) at the level of individual SNPs (maternal, pink and positive log2 FAB; paternal, blue and negative log2 FAB) reveal the existence of distinct maternal and paternal transcripts at the *H13* locus. The long *H13* transcripts are associated with maternally biased SNP sites and the short transcripts are associated with paternally biased SNP sites.

(**C** and **D**) Bar plots of the RNASeq data for *Gnas* in the ARN, DRN, liver and muscle demonstrate that a paternal bias exists at the gene level in the brain, but not the liver or muscle and expression is low in these tissues (C). Plots of the log2 FAB at the level of individual SNPs identify distinct transcripts from the maternal and paternal alleles (D). A maternal bias is detected for the *NESP* transcript (red) in the ARN and liver, and the four *XLαs* transcripts (dark blue) are associated with a paternal bias in the ARN (and DRN, not shown). Imprinting effects are not detected for the short *GSαs* (black) and *Ex1A* (sky blue) transcripts in any tissue.

Figure S7 (related to Figure 7). RNAscope probe detection of rapidly processed introns in the nascent RNA reveals allelic expression in vivo and monoallelic subpopulations of cells for noncanonical imprinted genes.

(**A**) Gene models for *Ddc* isoforms. Introns are numbered based on a flattened gene model that includes all isoforms.

(**B** and **C**) The expression level of *Ddc* exons and introns in the DRN as determined by RNASeq in B6 mice (B, FPKM, fragments per kilobase per million reads). Introns that are retained exhibit relatively

high expression, while introns that are rapidly processed exhibit low expression in the poly(A) captured mRNA. (C) The intron retention score for each intron is based on the ratio of intron expression level to the mean expression level of the flanking exons. This approach clearly distinguishes *Ddc* introns with high and low retention, including introns 6 and 19, respectively.

(**D** and **E**) In situ staining in the DRN for RNAscope probes targeting the retained *Ddc* intron 6 (D) and the rapidly processed intron 19 (E). The images demonstrate that probes targeting retained introns have a speckled staining pattern and reveal that RNA molecules that retain the intron are distributed throughout the nucleus (D). In contrast, probes targeting rapidly processed introns clearly reveal the focal site(s) of transcription in the nucleus due the fact that the RNA containing the rapidly processed intron is located at the site of transcription (E). Both biallelic (orange arrows) and monoallelic (green arrows) cells are clearly visible in the DRN for *Ddc*.

(**F** and **G**) Quantification of the number of partial nuclei in 14 micron cryosections of the ARN and DRN determined from confocal z-stacks of DAPI stained nuclei. (F) Examples of partial (yellow arrows) and whole (pink arrows) nuclei in confocal Z-stack images. (G) Counts of partial and whole nuclei.

(**H** and **I**) Fluorescent in situ hybridization detecting the nascent RNA for the noncanonical imprinted genes *Ahi1* (H) and *Ago2* (I). Confocal images reveal subpopulations of biallelic (two transcriptional foci) and monoallelic (one transcriptional site) cells for each gene. Examples of whole nuclei exhibiting monoallelic expression are indicated by the white arrow in the xy and xz planes. These results indicate that monoallelic subpopulations exist and are not due to sectioning artifacts.

Supplementary Results

Sensitive methods to detect canonical and noncanonical imprinting effects

Here, we detail our approach to accurately identify maternal and paternal canonical and non-canonical imprinting effects. For each tissue, we performed 8-9 biological replicates for each cross. To reduce variance related to dissections and to maintain library complexity, we pooled RNA equally from 4-6 animals for each biological replicate and constructed libraries from ~2-3 µg of total RNA. This approach reduces technical variance related to the microdissection of the tissues (particularly the ARN and DRN), provides additional statistical power by including more animals in the analysis, and increases the complexity of the starting library, since a single animal only yields ~0.5 µg of RNA from the ARN or DRN. We then performed deep sequencing, generating 80-100 million 59 bp single-end reads per replicate (Table S1). The acquisition of large numbers of reads (increased sampling) was performed to accurately assess allele-specific expression effects for genes with lower expression levels. In total, we generated 68 RNASeq datasets with this approach to analyze imprinting effects in the ARN, DRN, liver and muscle.

To develop a sensitive and accurate approach, we first set out to define SNP sites that accurately discriminate the expression of Cast and B6 alleles for our study. SNPs and indels that differentiate the Cast and B6 mouse strains have been publically annotated based on genome sequence data. However, technical effects related to library prep and read alignment, and biological effects related to genetic drift and RNA editing, could impact the fidelity of these sites for the measurement of allele-specific expression levels in our RNASeq data. Therefore, to first define a set of high fidelity SNP sites for our study, we performed RNASeq on the ARN, DRN, liver and muscle from the Cast and B6 mothers and fathers used to generate the F1i and F1r offspring in our breeding colony. We aligned the B6 parent transcriptome data to the mm9 reference genome and used

methods implemented in SamTools to define polymorphic sites (Li, 2011b; 2011a; Li et al., 2009). This analysis defined 86,335 sites that were heterozygous or homozygous at the transcriptome level relative to the mm9 annotated genome sequence in our B6 breeders. All of these variant sites were omitted from our study. Next, we aligned the Cast parent transcriptome data to the mm9 genome sequence and identified over 1 million heterozygous or homozygous sites. We omitted all of the heterozygous sites from our study, as well as all indels and SNP sites within 59 bp (one read length) of an indel. We then focused further analysis on the remaining homozygous SNP sites.

SamTools computes a statistical quality score for each variant site. By enforcing a stringent SNP quality score cutoff of 60, we constructed a table of 400,820 SNP sites that permit an analysis of allele-specific expression effects in our data. Greater than 99% of these SNPs are present in the publically annotated variant files for Cast and B6 genomes, which confirms the fidelity of the SNP sites between the genome and transcriptome data. For this SNP set, we found that 93% of SNP sites located greater than 59bp (one read length) apart from each other in the same imprinted gene (1 % FDR, see below) agree in terms of the direction of the parental bias (based on ARN, Fig. S1B). The mean number of SNPs per imprinted gene was 11 at this cutoff (Fig. S1B) and we are able to evaluate imprinting for 68% of genes expressed in the ARN, 77% in the DRN, 85% in the muscle and 82% in the liver (Fig. S1C). In total, we are able to analyze imprinting effects for 15,496 genes in the mouse genome, 91% of which are expressed in all four tissues (Fig. S1D). Thus, our SNP set yields excellent coverage and reproducibility between SNPs in the same gene for imprinting effects. We further found that this SNP set slightly improved the sensitivity to statistically detect imprinting effects in the data at a stringent 1% FDR compared to lower quality score cutoffs (Fig. S1E). The same finalized reference SNP set was used in our analysis of imprinting in each tissue type, below.

We next developed a method to evaluate our SNP set based on internal controls for noncanonical imprinting effects. Maternally-inherited mitochondrial genes or known canonical

imprinted genes could serve as internal controls, but these genes exhibit complete allele silencing and are therefore not representative of noncanonical imprinting effects. Previous studies have demonstrated that parent-of-origin effects influence X-inactivation (Chadwick and Willard, 2005; Fowlis et al., 1991) and that a bias exists to express the maternally inherited X chromosome (Xm) in female mice (Calaway et al., 2013; Gregg et al., 2010a; Wang et al., 2010). We reasoned that this maternal X chromosome bias could serve as a powerful internal control to evaluate the detection of noncanonical imprinting effects involving an allele bias. We determined that between 80-90% of ~500 X-linked genes analyzed in our study exhibit a maternal bias in each of the four tissue types (Fig. S2A). Histogram plots of the distribution of p-values for imprinting effects for X-linked genes indicate that the Xm bias is statistically detected in our data for most genes (Fig. S2B, and see below for statistical methods). Thus, by evaluating the proportion of maternally biased X-linked genes that are discovered at specific p-value cutoffs we can gain insights into whether different parameters, such as the SNP quality score improve the signal-to-noise in our data by increasing or decreasing the proportion of maternally biased X-linked genes detected. Importantly, we found that our SNP set yields improves detection of the maternal bias for X-linked genes compared to SNP quality score cutoffs less than 60 (Fig. S1F). Increasing the stringency of the quality score to cutoffs higher than 60 did not improve the signal:noise characteristics of the data at 1%, 2.5% and 5% controlled FDRs (Fig. S1F).

For the alignment of the F1i and F1r hybrid offspring RNASeq data, we used Novoalign, which allows for the incorporation of IUPAC base codes at polymorphic sites in the reference genome sequence, thereby reducing alignment biases and allowing for the utilization of a synthetic genome that incorporates both Cast and B6 sequence information. This alignment strategy improved the sensitivity to detect imprinting effects by ~11% compared to aligning to the mm9 reference genome sequence. The sensitivity to detect imprinting was further increased by omitting regions containing indels, which cannot be accommodated by IUPAC codes (Fig. S1G). While *improving the sensitivity*

to detect imprinting effects compared to aligning to the B6 genome, this alignment strategy did not change the proportion of maternally biased X-linked genes detected as a fraction of all imprinted genes detected at different p-value cutoffs, indicating that the *specificity is largely unchanged* by these different alignment strategies (not shown). Our analysis only included reads with unique alignments and excluded reads with multiple alignments.

To quantify allele-specific expression and imprinting effects, we generated a custom analysis pipeline that tallies the total number of Cast and B6 allele reads for each ensemble-annotated gene based on our SNP set. Due to the fact that a single read can cross multiple SNP sites, our approach was adjusted to ensure that each read is tallied only once and we summed together the tallies across all SNPs for a given gene to arrive at a gene-level analysis. To statistically define imprinting effects across F1i and F1r biological replicates for each tissue, we used a generalized linear modeling (glm) approach implemented in the statistical package edgeR (McCarthy et al., 2012; Robinson et al., 2010). Our approach tests for a main effect of parent by contrasting the full model: (allele counts ~ sample (1,2,3…18) + strain (Cast versus B6) + parent (maternal versus paternal)) to the reduced model (allele counts ~ sample + strain). Thus, the approach absorbs variance in the data related to sample effects and strain-specific allelic effects. This is important since most autosomal and X-linked genes exhibit a Cast or B6 allele bias due to genetic strain effects. We found that this glm implementation in edgeR had increased sensitivity and specificity for the detection of imprinting effects compared to other statistical methods we tested. In this approach, normalization factors for library size are computed from the entire raw read count dataset and then applied to the allele count dataset in the DGEList object. To estimate type I statistical errors (FDR) in our data at different p value cutoffs, we devised an FDR estimation approach that involves randomly permuting the allele-specific expression data. In our approach, we intermingle data from F1i and F1r samples in each permutation, which has the effect of scrambling the parental origin of the samples relative to the glm design matrix (Fig. S1H). For each permutation, we calculate the number of false positives at

different p-value cutoffs, which, in our study, reaches a plateau after ~500 iterations. Therefore, for each dataset, we used this approach to calculate the mean number of false positives detected at each p-value cutoff and compared this to the actual number of genes detected in order to estimate the FDR (false / actual) (results presented in Fig. 1 in the main text). In practice, this approach yields results that are slightly more conservative than the Benjamini-Hochberg (BH) correction method for the DRN, liver and muscle datasets and similar to the BH method for the ARN dataset.

Finally, we devised a method to estimate type II statistical errors (false negatives) in our data. In the past, studies have used known canonical imprinted genes (Wang et al., 2008) or X-linked genes in males as internal positive controls (Gregg et al., 2010b). However, these effects involve allele-specific silencing, and are therefore not appropriate internal controls for the detection of noncanonical imprinting effects. As introduced above, our approach takes advantage of the Xm expression bias that occurs in female mice and is clearly detected in the ARN, DRN, muscle and liver in our study (Fig. S2A and B). Thus, we are able to gain insights into the type II statistical errors at different p value cutoffs by determining the proportion of maternally biased X-linked genes that are detected. For example, in the ARN at a controlled FDR of 1%, we detect ~35% of the expressed and maternally biased X-linked genes in this tissue (Fig. S2C). Therefore, and as detailed in the main text, these internal controls suggest that our screen for noncanonical imprinting effects is not saturated. Notably, our approach relies on SNPs to distinguish allelic expression, and therefore, genes without SNPs or with low expression are especially likely to be missed.

Noncanonical imprinting effects can arise independently from canonical imprinting

(continued from main manuscript)

We identified noncanonical imprinting effects influencing three genes near the known imprinted gene, *Inpp5f* (Fig. S3H). The *Inpp5f* locus gives rise to distinct isoforms from the maternal and paternal alleles in a tissue-specific manner (Wood et al., 2007a), and we explore imprinting at the transcript

level in a separate section in the main manuscript. In our gene level analysis, *Inpp5f* exhibits a paternal bias in the ARN and DRN, but not the liver or muscle (Fig. S3I). Similar to *Plagl1,* the *Inpp5f* locus is not considered part of an imprinted gene cluster. However, we found that the neighboring genes, *Bag3, BC017158* and *Tgfb1i1* exhibit a tissue-specific maternal allele expression bias. *Bag3* and *BC017158* exhibit a significant maternal bias in the ARN, but not the DRN, liver or muscle (Fig. S3J and K). On the other hand, *Tgfb1i1,* which is the most distant gene from *Inpp5f* in the cluster, exhibits a significant maternal bias in both the ARN and DRN, but not the liver or muscle (Fig. S3L). Pyrosequencing confirmed that *Tgfb1i1* exhibits a significant maternal bias in the ARN, but not in the liver, in both Cast x B6 (Fig. S3M) and PWD/J x A/J hybrid mice (Fig. S3N). In conclusion, our analysis reveals the existence of highly tissue-specific non-canonical imprinting effects that arise near canonical imprinted genes. Interestingly, the tissue-specific imprinting patterns exhibited by these genes are different from the imprinting pattern exhibited by the neighboring canonical imprinted genes.

Noncanonical imprinted genes exhibit allele-specific expression effects in discrete subpopulations of neurons (continued from main manuscript)

Noncanonical imprinting effects involving maternal or paternal allele expression biases could be due to distinct, but overlapping transcripts from maternal versus paternal alleles (Fig. 6A). Genes that are known to exhibit these effects include *Gnas, Mest, H13, Herc3, Commd1, Trappc9, Inpp5f* and *Blcap* (Bastepe, 2007; Gregg et al., 2010b; McCole and Oakey, 2008; Nakabayashi et al., 2002; Wood et al., 2007b; 2008). In our RNASeq analysis, all of these genes exhibit a noncanonical imprinting pattern. However, we previously found that analyzing the data at the level of individual SNPs, rather than at the gene level, can reveal the existence of distinct maternal and paternal allele transcripts (Gregg et al., 2010b). For example, *H13* and *Gnas* exhibit an allele expression bias at the gene level that can be clearly resolved as distinct maternal versus paternal transcripts at the SNP level (Fig. S6A-D). To test whether overlapping transcripts are a common feature of noncanonical imprinted

genes, we devised an approach to identify isoform-specific imprinting effects by RNASeq in each tissue. We flattened the ensembl annotated transcript models into one gene model that included all exons for each gene. Next, we summed reads for SNPs located within the same exon and constructed a data table composed of exon level allele-specific expression data for each gene model. We then modified the glm function in the edgeR package to test for an interaction effect between exon expression level and parental origin of the allele to identify genes with exons that significantly disagree in terms of their expression level from the maternal versus paternal alleles. P-values were adjusted using the Benjamini-Hochberg approach. With this approach, we determined that *H13, Commd1, Trappc9, Herc3, Inpp5f, Blcap, Mest, Ube3a* and *Gnas* are the only genes with overlapping, allele-specific isoforms (BH adjusted p-value < 0.01). Therefore, our approach robustly detects all known cases of isoform-specific imprinting and our results reveal that most noncanonical imprinting effects are not due to this phenomenon.

References

Bastepe, M. (2007). The GNAS Locus: Quintessential Complex Gene Encoding Gsalpha, XLalphas, and other Imprinted Transcripts. Curr. Genomics *8*, 398–414.

Calaway, J.D., Lenarcic, A.B., Didion, J.P., Wang, J.R., Searle, J.B., McMillan, L., Valdar, W., and Pardo-Manuel de Villena, F. (2013). Genetic Architecture of Skewed X Inactivation in the Laboratory Mouse. PLoS Genet *9*, e1003853.

Chadwick, L.H., and Willard, H.F. (2005). Genetic and parent-of-origin influences on X chromosome choice in Xce heterozygous mice. Mamm Genome *16*, 691–699.

Fowlis, D.J., Ansell, J.D., and Micklem, H.S. (1991). Further evidence for the importance of parental source of the Xce allele in X chromosome inactivation. Genet Res *58*, 63–65.

Gregg, C., Zhang, J., Butler, J.E., Haig, D., and Dulac, C. (2010a). Sex-specific parent-of-origin allelic expression in the mouse brain. Science *329*, 682–685.

Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G.P., Haig, D., and Dulac, C. (2010b). High-resolution analysis of parent-of-origin allelic expression in the mouse brain. Science *329*, 643–648.

Li, H. (2011a). Improving SNP discovery by base alignment quality. Bioinformatics *27*, 1157–1158.

Li, H. (2011b). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics *27*, 2987–2993.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res *40*, 4288–4297.

McCole, R.B., and Oakey, R.J. (2008). Unwitting hosts fall victim to imprinting. Epigenetics *3*, 258–260.

Miller, R.A., Harper, J.M., Dysko, R.C., Durkee, S.J., and Austad, S.N. (2002). Longer life spans and delayed maturation in wild-derived mice. Exp. Biol. Med. (Maywood) *227*, 500–508.

Nakabayashi, K., Bentley, L., Hitchins, M.P., Mitsuya, K., Meguro, M., Minagawa, S., Bamforth, J.S., Stanier, P., Preece, M., Weksberg, R., et al. (2002). Identification and characterization of an imprinted antisense RNA (MESTIT1) in the human MEST locus on chromosome 7q32. Hum Mol Genet *11*, 1743–1756.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Wang, X., Soloway, P.D., and Clark, A.G. (2010). Paternally biased X inactivation in mouse neonatal brain. Genome Biol *11*, R79.

Wang, X., Sun, Q., McGrath, S.D., Mardis, E.R., Soloway, P.D., and Clark, A.G. (2008). Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. PLoS ONE *3*, e3839.

Wood, A.J., Bourc'his, D., Bestor, T.H., and Oakey, R.J. (2007a). Allele-specific demethylation at an imprinted mammalian promoter. Nucleic Acids Res *35*, 7031–7039.

Wood, A.J., Roberts, R.G., Monk, D., Moore, G.E., Schulz, R., and Oakey, R.J. (2007b). A screen for retrotransposed imprinted genes reveals an association between X chromosome homology and maternal germ-line methylation. PLoS Genet *3*, e20.

Wood, A.J., Schulz, R., Woodfine, K., Koltowska, K., Beechey, C.V., Peters, J., Bourc'his, D., and Oakey, R.J. (2008). Regulation of alternative polyadenylation by genomic imprinting. Genes Dev *22*, 1141–1146.

Supplementary Methods

Animals

To generate F1 hybrid offspring, we mated C57BL/6J and CastEiJ males and females, as well as PWD/J and A/J males and females (Jackson Laboratories, Bar Harbor, Maine). Wild-derived, outbred mice captured from Idaho (Id mouse line) have been previously described (Miller et al., 2002). *Th* mutant mice were backcrossed onto a B6 background for at least 8 generations and were a kind gift from Dr. Richard Palmiter (University of Washington). Mice were weaned at P21 and grouped housed with same-sex littermates. Mice were maintained on a 12hr light/dark cycle (lights off at 6pm and lights on at 6am), and given water and food *ad libitum (*Harlan Tecklad rodent diet 2920X; Madison, WI.). All animal experiments were performed according to regulations at the University of Utah.

RNA Isolation and RNASeq

The ARN, DRN, liver and thigh muscle were microdissected from female F1 hybrid mice at 8-10 weeks of age. To analyze SNPs in the parents, these tissues were dissected from C57BL/6J and CastEiJ breeders at approximately 6 months of age following the birth of litters. The ARN dissection includes the ventral medial hypothalamus. The DRN dissection includes portions of the ventral periaqueductal grey. The tissues were stored in RNAeasy Kit lysis buffer (Qiagen) at -80C until the RNA was extracted using the RNAeasy Micro Kit (Qiagen). RNA was pooled from 4-5 daughters from different litters to provide ~3 ug of total RNA for each biological replicate. Samples were prepared for RNASeq using the TruSeq RNA sample preparation kit v2 (RS-122-2001, Illumina). Single-end sequencing of the libraries was performed using the HiSeq 2000 (Illumina). We performed nine biological replicates for each cross for the ARN and DRN studies and eight biological replicates for each cross for the liver and muscle studies. The methods for analyzing the data and descriptive statistics are presented in the text and supplementary information.

Allele-Specific chIP

Chromatin was isolated from the hypothalamus of Cast x B6 F1 hybrid mice and chromatin immunoprecipitation for H3K9ac and H3K9me3 was performed using the Imprint Chromatin Immunoprecipitation kit according to the manufacturer's instructions (CHP1-24RXN, Sigma-Aldrich). We used the following antibodies: mouse anti-H3K9ac (ab4441, Abcam) and rabbit anti-H3K9me3 (ab8898, Abcam).

mRNA preparation for pyrosequencing

Total RNA was purified from the ARN, DRN, liver or muscle from individual Cast x B6 F1 hybrid offspring using MicroElute Total RNA kit (R6831-02, Omega). The cDNA library was generated using the qScript cDNA supermix (P/N84034, Quanta) and oligo(dT) primers + random hexamer primers according to the manufacturer's instructions.

Pyrosequencing

Pyrosequencing to analyze allele-specific expression effects or allele-specific chIP for specific genes was performed using the following reagents: Pyromark Gold Q24 reagents (970802, Qiagen), Pyromark Q24 cartridge (979202, Qiagen), Pyromark Q24 plate (979201, Qiagen), Pyromark Binding Buffer (979002, Qiagen), Denaturation Solution (979007, Qiagen), Pyromark annealing buffer (979009, Qiagen), Pyromark wash buffer (979008, Qiagen), streptavidin-sepharose column (17-5113-01, GE). Primers were designed using Pyromark Assay Design AW software (PyroMarkQ24 2.0.6, Qiagen). Pyrosequencing was performed and analyzed using the PyroMark Q24 system and software (Qiagen). Amplification primers and sequencing primers are provided in Tables S7, S8 and S12 for mRNA and chIP studies, as indicated in the text. For each experiment, we identified SNP sites that discriminate Cast and B6 alleles and amplified ~100-150 base pairs surrounding the SNP site. The sequencing primer was designed upstream of the SNP site and the ratio of Cast versus B6 base calls

was determined in each sample according to the Pyromark software. We performed 4-8 biological replicates for each cross and calculated the Cast:B6 allele expression (mRNA) or enrichment (chIP) ratio for each replicate. Using a one-tailed t-test for RNASeq validation studies and a two-tailed t-test for chIP studies, we compared the Cast:B6 allele expression or enrichment ratio between the F1i and F1r hybrid offspring to test for statistically significant differences between the two crosses. Significant differences in the Cast:B6 allele expression or enrichment ratio between F1i and F1r offspring arise due to parent-of-origin effects influencing Cast and B6 alleles.

Wild-Mouse RNASeq Analysis

The hypothalamus was stored in RNAeasy Kit lysis buffer (Qiagen) at -80C until the RNA was extracted using the RNAeasy Micro Kit (Qiagen). Samples were prepared for RNASeq using the TruSeq RNA sample preparation kit v2 (RS-122-2001, Illumina). Single-end sequencing of the libraries was performed using the HiSeq 2000 (Illumina). We performed RNASeq to profile the transcriptome of the hypothalamus for each mother and father and aligned the data to the mm9 reference genome sequence in order to first identify homozygous and heterozygous SNP sites in the parents that could potentially distinguish maternally versus paternally inherited alleles in the daughters. Then, using custom Python code, we analyzed the transcriptome of the daughters and defined genes for which the parental origin of each allele could be unambiguously defined. In trios #1, #2 and #3, we identified a total of 11040, 10466 and 8804 homozygous and heterozygous SNPs, respectively, between the mothers and fathers (Phred scaled SNP Quality score >= 60). These SNP sites enabled us to distinguish allele-specific expression effects for 844, 850 and 718 genes for the daughters in trio #1, #2 and #3, respectively. The number of SNPs in each gene was variable between the different daughters and we summed reads across SNPs in the same gene to obtain a relative expression estimate for the maternal versus paternal alleles for each daughter.

RNAscope Staining and Nuclei Analysis

RNAscope probes targeting specific introns were designed by Advanced Cell Diagnostics (www.ACDbio.com) and staining was performed using ACD 2-plex RNAscope kits according to the manufacturers instructions on 14 micron tissue cryosections from B6 mice. Probes used in this study are available for ordering from ACD. To identify rapidly processed introns to define allelic transcriptional activity, we computed the expression level of each exon and intron for each gene from our RNASeq data and calculated an intron retention score (IRS). For each gene and tissue, we defined relatively retained (high IRS) versus rapidly processed introns (low IRS). By targeting rapidly processed introns with RNAscope probes we can resolve allele expression at the cellular level for any gene.

Female B6 adult whole brains were fresh frozen in OTC and stored at -80ºC and later sectioned at 14µm on a cryostat. Sections were allowed to dry at RT and then stored at -80ºC. For nascent RNA detection sections were treated according to RNAscope Sample Preparation and Pretreatment Guide for Fresh Frozen Tissue (Part 1) and RNAscope 2.0 2-Plex Detection Kit user manuals (www.ACDBio.com). Images were taken on Zeiss ApoTome2 microscope with a 40x oil objective. For the ARN, two images were taken from three different sections for a total of 6 images per probe for each experiment. For the DRN, three images were taken from three different sections for a total of 9 images per probe for each experiment. The number of nuclei with RNA probe signal was counted and from these the percentage of nuclei with a single or double dot was calculated. The average of the percentage of single dot cells was determined for each animal and brain region.

To analyze whole versus partial nuclei mice were transcardially perfused with 4% PFA. The brain was dissected and fixed for 6hrs at RT in 4% PFA. The brain was immersed in 10% sucrose in 1xPBS at 4ºC until it sank to the bottom of the container. This step was repeated with 20% and 30% sucrose. The brain was then embedded with OTC and stored at -80ºC. ARN and DRN 14µm sections were collected using a cryostat. The sections were allowed to dry and then stored at -80ºC. RNA signal was detected following the RNAscope Chromogenic Assay Sample Preparation for Fixed Frozen Tissue (Part 1) technical note and RNAscope 2-Plex Kit (RED) (Part 2) user manual with the

following change. The sections were allowed to dry shortly after detection of the signal using FAST RED A&B. Then, the slides were mounted using Prolong Gold antifade reagent with DAPI. Tissue sections were imaged with a Zeiss ApoTome2 microscope using a 60x oil objective. Z-stack images were taken through the entire thickness of the section. Z-stacks were rotated to visualize xz planes. Ten images were taken per section from at least 3 sections per animal from ARN and DRN and Z-stacks were rotated to visualize xz planes. For quantification of whole and partial nuclei, 11 xz planes were selected (from 1 to 1000 planes we selected every 100th plane) from each image for counting. The total number of whole nuclei and cut nuclei from all 110 images was counted for ARN and DRN and the percentage of cut nuclei was calculated.

Intron Retention Analysis

Using custom python and R scripts we computed the FPKM for each exon and intron for each gene in the ARN, DRN, liver and muscle. We calculated an intron retention score for each intron as the intron FPKM divided by the mean FPKM for the two flanking exons.

Behavior Studies

Mutant mice with a heterozygous deletion of the first exon of tyrosine-hydroxylase (*Th*) on the C57Bl/6J background strain were obtained from the lab of Richard Palmitter, University of Washington. Heterozygous *Th* females and males were paired with wild-type males and females, respectively, and bred at the University of Utah to produce heterozygous offspring with either a maternal *Th* mutation ($Th^{-/+}$) or a paternal mutation ($Th^{+/-}$), and wild-type littermates.

Open Field

Between 7-9 weeks of age $Th^{-/+}$ and $Th^{+/-}$, and their wild-type littermates, were tested for exploratory behavior in the open-field test. Mice were habituated to the testing room for at least one hour, and tested during the light portion of the day (between 11:30am to 4:30pm) in 40cm square open field

boxes (1,600cm^2 area), under dim, 40 lux illumination. Behavior in the open field was video recorded, and the center-point of the mouse was tracked with Noldus EthoVision XT software (version 9.0.726). EthoVison was used to calibrate the size of the open field arena and to outline the center and wall zones on the video recordings. A 20cm square box was drawn to define the center zone (400cm^2 area) of the open field, with each side equidistant (10cm) from one of the four 40cm high walls; the wall zone (1,200cm^2 area) was the surrounding area between the center zone and the walls of the open field box. To start the test, mice were placed into the wall zone and EthoVision started tracking the mice with a one second delay, and continued for 15 min. Total time spent in the center-zone of the arena (in sec.), and total distance traveled in the arena (in cm) was analyzed from the tracking data. Statistical differences were measured with two-way ANOVAs using genotype and sex as main factors using Prism 5 for Mac OS X software (GraphPad Software Inc.). The number of animals was 11 for each sex by genotype group, except for wt males from heterozygous fathers where n=10.

Sucrose Intake

Mice were tested for sucrose intake at least one week after open field tests, and between 8 and 11 weeks of age. At 9am immediately before the first day of testing, body weight was measured and mice were individually housed into clean cages. The mice were fed *ad libitum* during testing, and given a choice of drinking either a sucrose solution or water from two 50 ml conical tubes plugged at the opening with stoppers fitted with sippers. The two tubes were tipped over, and lowered side-by-side so that the sippers pointed down into the cage through the wire-top feeding rack. Each mouse was tested for intake of water compared to a series of 0%, 0.32%, 1%, or 3.2% sucrose solution in drinking water for two days at each sucrose concentration, for a total of eight days of testing. On the first two days, both tubes were filled with drinking water (0% sucrose) to acclimate to the task and control for tube preference. Between days one and two, the positions of the two tubes were swapped to control for a side preference. On days 3 and 4, the tubes were replaced with one tube of fresh water and one tube of 0.32% sucrose to test sucrose intake; the position of the water and sucrose

tube was switched between day 3 and 4. Days 5 and 6 tested with 1% sucrose, and days 7 and 8 tested with 3.2% sucrose, switching the side of the water and sucrose tube on alternate days. Sucrose and water intake was measured twice daily at 9am and 6pm to obtain measurements for both intakes during the light portion and the dark portion of the day, and the average daily intake was calculated for each two-day intake at a particular concentration. Sucrose and water intake were normalized to body weight and statistical differences were calculated with a repeated measures ANOVA using genotype as the between factor and sucrose concentration as the within factor in Prism software. The number of animals per group was the following: $Th^{-/+}$ n=12 (6 male and 6 female), wt littermates n=12 (6 male and 6 female); and $Th^{+/-}$, n=13 (7 male and 6 female), wt littermates n=16 (8 male and 8 female).

Supplementary Tables

Table S1 (related to Experimental Procedures). Uniquely mapped read counts for RNASeq replicates for the ARN, DRN, muscle and liver experiments. Reads were mapped using NovoAlign.

Table S2 (related to Figure 1). EdgeR generalized linear model (glm) results for imprinting effects, permutation test false discovery rate (FDR) calculations and the proportion of maternally biased X-linked genes detected. 'Imprinting' labeled tabs provide glm results for imprinting effects for the ARN, DRN, liver and muscle. The 1% FDR cutoff point is highlighted in yellow.

Column Definitions: ID, ensemble gene ID; logFC, log2 fold allele bias (negative is paternal bias, positive is maternal bias); logCMP, log2 counts per million; LR, likelihood ratio; PValue, p-value for a main effect of the parental origin of the allele (imprinting); Benjamini-Hochberg adjusted p-value; chr, chromosome; name, gene name.

'FDR' labeled tabs provide the results of the permutation test and present the mean number of false imprinted genes detected after 500 iterations at each p-value cutoff for the ARN, DRN, liver and muscle data. The fraction of maternally biased X-linked genes detected is also presented as a measure of the sensitivity to detect non-canonical imprinting. The p-value at which a 1% FDR is detected is highlighted in yellow.

Column Definitions: pValue, the p-value cutoff for imprinting; Genes_Examined, total number of genes expressed with SNPs in the dataset; Real_Imprinted_AutosomesANDchrX, total number of imprinted genes detected at the indicated p-value cutoff; FDR_mean, mean number of false positives detected at the p-value cutoff; FDR_std, standard deviation of the number of false positives detected at the p-value cutoff; FalseDiscoveryRate, estimated false

discovery rate at each p-value cutoff; FractionOfMaternalChrXGenesDetected, the proportion of maternally biased X-linked genes detected at each p-value cutoff.

Table S3 (related to Figure 1). Summary of the autosomal imprinted genes detected in our study at the 1% FDR cutoff. The table provides information on the tissue in which the imprinting was detected, maternal or paternal expression, canonical or non-canonical imprinting, clustering, transcript level imprinting effects, and novelty, as well as other metadata.

Column Definitions: ID, ensemble gene ID; LogFC, log2 fold allele bias (negative is paternal bias, positive is maternal bias); PValue, p-value for imprinting (NAs are placed if the p-value is below the 1% FDR cutoff); Canonical, true indicates canonical imprinting and false indicates non-canonical imprinting; Transcript_Specific_Imprinting, true/false indicates whether distinct transcripts were detected from the maternal and paternal alleles; external_gene_id, gene name; start_position, genomic transcriptional start site; end_position, genomic transcriptional end site; gene_biotype, gene classification; Chr, chromosome; Clustered, true/false indicates clustering with other imprinted genes located within 1Mb; Imprinted, true/false indicates significant imprinting in each tissue type; Known, true/false indicates whether the imprinted gene has been previously identified.

Table S4 (related to Figure 3). Identification of canonical and non-canonical imprinted genes located in gene clusters and in novel genomic regions. Clustered genes are < 1Mb apart in the genome.

Column Definitions: ID, ensemble gene ID; LogFC, log2 fold allele bias (negative is paternal bias, positive is maternal bias); PValue, p-value for imprinting (NAs are placed if the p-value is below the 1% FDR cutoff); Canonical, true indicates canonical imprinting and false indicates non-canonical imprinting; external_gene_id, gene name; start_position, genomic transcriptional start site; end_position, genomic transcriptional end site; gene_biotype, gene

classification; Chr, chromosome; ClusterID, genes with the same number are part of the same cluster; KnownCluster, true/false indicates whether the gene cluster is known (true) or novel (false).

Table S5 (related to Figure 3). Summary of pyrosequencing validations in Cast x B6 and PWD/J x A/J hybrid offspring for the ARN, DRN, muscle and liver. The 'Cast x B6' tab provides a summary of validations performed in Cast x B6 hybrid mice. ARN data is in orange columns, DRN data is in green columns, muscle data is in brown columns and liver data is in blue columns.

Column Definitions: GeneNames, imprinted gene names; Expressed, true/false indicates whether the gene is expressed in a particular tissue; Imprinted, true/false indicates whether the gene is significantly imprinted in a particular tissue according to RNASeq; Hybrid_Cross, indicates Cast x B6 hybrid mice are analyzed for the pyrosequencing; RNASeq_Log2FoldAlleleBias, maternal (positive) and paternal (negative) allele bias based on RNASeq; RNASeq_Pvalue, p-value for imprinting; PyroResult_AlleleBias, maternal or paternal allele bias direction detected by pyrosequencing (NA indicates the gene was not tested); Pyro_Pvalue, the one tailed t-test p-value indicates whether imprinting ($P < 0.05$) is detected by pyrosequencing; Pyro_NumberOfReplicates, indicates the number of F1i and F1r cross biological replicates used in the pyrosequencing experiment; Pyro_RNASeq_Agreement, true/false indicates whether the RNASeq and pyrosequencing results agree; SNP_Chr_Position, indicates the mm9 genomic coordinates for the SNP. Pyrosequencing primer sequences for each gene are provided in the table.

The 'PWD'J x A/J' tab provides a summary of validations performed in PWD/J X A/J hybrid offspring.

Column Definitions: Gene, imprinted gene names; Chromosome, the chromosome location of

the gene; TissueAnalyzed, indicates whether the pyrosequencing was performed on ARN, DRN, liver or muscle; RNASeq Result, indicates the imprinting effect detected by RNASeq; HybridCross, PWD/J x A/J reciprocal hybrids; PyroSeq Result, indicates the direction of the allele bias detected by pyrosequencing; PyroSeq P Value, indicates the one tailed t-test p-value result for an imprinting effect by pyrosequencing ($P<0.05$); PyroSeq Replicates, number of biological replicates for the F1i and F1r cross for each experiment; SNP_site, chromosomal coordinates for the SNP site location. Pyrosequencing primer sequences are provided for each gene.

Table S6 (related to Figure 3). Summary of data for the analysis of noncanonical imprinted genes in wild-derived mouse trios.

Table S7 (related to Figure 5). Summary of allele-specific chromatin immunoprecipitation pyrosequencing experiments in the hypothalamus of adult female Cast x B6 F1 hybrid offspring.

Column Definitions: Type, Canonical or non-canonical imprinted genes; Gene, gene name; Preferentially Expressed Allele, the preferentially expressed allele identified by RNASeq; Tissue, the tissue from which the ChIP was performed; One-tailed t-test H3K9ac (enriched allele), p-value for allele-specific enrichment of H3K9ac and the enriched allele is indicated in brackets; One-tailed t-test H3K9me3 (enriched allele), p-value for allele-specific enrichment of H3K9me3 and the enriched allele is indicated in brackets; TSS location, genomic location of the transcriptional start site; SNP Site, the location of the targeted SNP site; distance from TSS, the distance the SNP is located from the transcriptional start site; Base Call, the B6/Cast base calls at the SNP site; SNP position, location of the SNP relative to the promoter and gene body. Primer sequences for the pyrosequencing analysis.

CHAPTER 3

DIVERSE NONGENETIC, ALLELE-SPECIFIC EXPRESSION EFFECTS SHAPE GENETIC ARCHITECTURE AT THE CELLULAR LEVEL IN THE MAMMALIAN BRAIN

The following chapter is reprinted with permission from Cell Press.

Huang, W. C., Ferris, E., Cheng, T., Horndli, C. S., Gleason, K., Tamminga, C., . . . Gregg, C. (2017). Diverse non-genetic, allele-specific expression effects shape Genetic architecture at the cellular level in the mammalian brain. *Neuron, 93*(5), 1094-1109 e1097.

Article

Neuron

Diverse Non-genetic, Allele-Specific Expression Effects Shape Genetic Architecture at the Cellular Level in the Mammalian Brain

Highlights

- In vivo genome-wide screen uncovers diverse non-genetic allelic effects
- Non-genetic allelic effects are prevalent in the neonatal mouse brain
- Allelic effects cause mosaics of mutant and wild-type cells for heterozygous mutations
- Allelic effects exist in the primate brain and impact genes linked to mental illness

Authors

Wei-Chao Huang, Elliott Ferris, Tong Cheng, ..., Kenneth M. Boucher, Jan L. Christian, Christopher Gregg

Correspondence

chris.gregg@neuro.utah.edu

In Brief

Huang and Ferris et al. uncover diverse forms of non-genetic allelic effects in vivo in the mouse and primate brain that can interact with heterozygous mutations to generate mosaics of brain cells that differentially express mutant versus wild-type alleles.

Huang et al., 2017, Neuron 93, 1094–1109
March 8, 2017 Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.neuron.2017.01.033

CellPress

CellPress

Neuron
Article

Diverse Non-genetic, Allele-Specific Expression Effects Shape Genetic Architecture at the Cellular Level in the Mammalian Brain

Wei-Chao Huang,[2,9] Elliott Ferris,[2,9] Tong Cheng,[2] Cornelia Stacher Hörndli,[2] Kelly Gleason,[7] Carol Tamminga,[7] Janice D. Wagner,[8] Kenneth M. Boucher,[4,5,6] Jan L. Christian,[2,4] and Christopher Gregg[1,2,3,10,*]
[1]Robertson Neuroscience Investigator, New York Stem Cell Foundation
[2]Departments of Neurobiology & Anatomy
[3]Department of Human Genetics
[4]Department of Internal Medicine
[5]Cancer Biostatistics Shared Resource
[6]Huntsman Cancer Institute
University of Utah School of Medicine, Salt Lake City, UT 84112, USA
[7]Department of Psychiatry, UT Southwestern, Dallas, TX 75390-9127, USA
[8]Department of Pathology, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA
[9]Co-first author
[10]Lead Contact
*Correspondence: chris.gregg@neuro.utah.edu
http://dx.doi.org/10.1016/j.neuron.2017.01.033

SUMMARY

Interactions between genetic and epigenetic effects shape brain function, behavior, and the risk for mental illness. Random X inactivation and genomic imprinting are epigenetic allelic effects that are well known to influence genetic architecture and disease risk. Less is known about the nature, prevalence, and conservation of other potential epigenetic allelic effects in vivo in the mouse and primate brain. Here we devise genomics, in situ hybridization, and mouse genetics strategies to uncover diverse allelic effects in the brain that are not caused by imprinting or genetic variation. We found allelic effects that are developmental stage and cell type specific, that are prevalent in the neonatal brain, and that cause mosaics of monoallelic brain cells that differentially express wild-type and mutant alleles for heterozygous mutations. Finally, we show that diverse non-genetic allelic effects that impact mental illness risk genes exist in the macaque and human brain. Our findings have potential implications for mammalian brain genetics.

INTRODUCTION

Recent genomic studies of neuropsychiatric disorders created a wealth of data on the genetics of these disorders (Gratten et al., 2014; McCarroll et al., 2014). Less is known about how epigenetic mechanisms interface with genetic mutations to cause brain dysfunction. Studies of genomic imprinting and random X inactivation demonstrated that epigenetic effects impacting a single allele can profoundly influence genetic architecture, phenotypes, and disease susceptibility (Deng et al., 2014a; Peters, 2014). Genomic imprinting effects are relatively enriched in the brain, but they impact the expression of fewer than 200 autosomal genes in the mouse and human (Babak et al., 2015; Bonthuis et al., 2015; Perez et al., 2015). Thus, the mechanisms controlling gene expression for most autosomal genes are thought to regulate both alleles equally. However, since genetic risk factors for mental illness are frequently heterozygous in affected individuals—meaning only one allele is mutated—the discovery of other epigenetic allelic effects in vivo that influence the expression of wild-type (WT) versus mutant (MT) alleles could improve our understanding of brain genetics.

Autosomal, epigenetic allele-specific expression (ASE) effects other than imprinting have been described (Chess, 2016). In vivo, antigen receptors, olfactory receptors (ORs), and clustered protocadherins exhibit monoallelic expression. From in vitro studies, random monoallelic effects have also been observed for many autosomal genes in human and mouse lymphoblastoid cell lines (Gimelbrant et al., 2007; Zwemer et al., 2012), neural stem cell lines (Jeffries et al., 2012), and embryonic stem cell (ESC) lines (Eckersley-Maslin et al., 2014; Gendrel et al., 2014). Further, studies of human ESCs showed that ASE and allele-specific chromatin structures are widespread (Dixon et al., 2015). However, these studies focused on cell lines, which can exhibit epigenetic instability that impacts allelic expression (Mekhoubad et al., 2012; Nazor et al., 2012; Stadtfeld et al., 2012).

Studies of transcription at the single-cell level also uncovered autosomal ASE effects (Borel et al., 2015; Deng et al., 2014b; Marinov et al., 2014; Raj and van Oudenaarden, 2008), though it is unclear which effects are due to transcriptional noise and which are bona fide in vivo ASE effects. A recent single-cell transcriptome analysis of clonally derived mouse fibroblasts and human T cells concluded that clonal, random monoallelic effects similar to X inactivation are rare on the autosomes

(Reinius et al., 2016); this challenges previous studies of random monoallelic effects in cell lines. Overall, a better understanding of the nature, diversity, prevalence, and conservation of epigenetic ASE effects in vivo is needed.

ASE effects in vivo in the mouse (Crowley et al., 2015; Pinter et al., 2015) and in different human tissues (Leung et al., 2015; Roadmap Epigenomics Consortium et al., 2015) have been largely attributed to genetic variation in *cis-regulatory* regions; this can cause allelic differences in chromatin states and gene expression (Heinz et al., 2013; Kasowski et al., 2013; Kilpinen et al., 2013). Currently, in vivo approaches to detect epigenetic random monoallelic effects are limited to an indirect chromatin signature derived from cell lines (Nag et al., 2013; Savova et al., 2016). Thus, beyond a few select cases, we know little about the nature and prevalence of non-genetic ASE effects in vivo.

Here, we introduce a genomics strategy and statistical framework to perform genome-wide screens for diverse forms of non-genetic allelic expression effects in vivo in the mouse and primate brain. The approach is designed to detect imprinting, random monoallelic expression and other possible allelic effects. We apply our methodology in the mouse to investigate whether non-genetic ASE effects are especially prevalent for specific developmental stages, brain regions, and tissue types and whether they impact the cellular expression of heterozygous mutations in vivo. By further screening for allelic effects in the macaque brain, we investigate the conservation of non-genetic allelic effects between mice and primates and determine whether these effects impact genes linked to mental illness in the macaque and human brain. Our study provides evidence for diverse non-genetic allelic effects in the mammalian brain that are not due to imprinting but can shape genetic architecture at the cellular level and arise in a developmental-stage and cell-specific manner.

RESULTS

An In Vivo Genome-wide Approach to Detect Diverse Non-genetic ASE Effects

To screen for diverse ASEs in vivo, we developed a genomics approach that tests the assumption that most genes express their maternal and paternal alleles equally (allele co-expression). The approach is based, in part, on our previous study of genomic imprinting (Bonthuis et al., 2015) in which RNA-seq is performed on tissue from CastEiJ (Cast) × C57BL6/J (B6) F1 hybrid offspring derived from initial (F1cb, Cast mother × B6 father) and reciprocal (F1bc, Cast father × B6 mother) crosses. Here, instead of defining imprinting effects, we test for allele co-expression by examining the correlation of maternal and paternal allele expression levels across many RNA-seq biological replicates from different F1cb and F1bc hybrid mice. The approach potentially detects: (1) allele co-expression effects (CoEEs), whereby maternal and paternal allele expression is strongly correlated; (2) differential allele expression effects (DAEEs), whereby the maternal and paternal alleles are weakly correlated or not correlated; and (3) antagonistic allele expression effects (AAEEs), for which increased expression of one allele is associated with decreased expression of the other (Figure 1A). At the cellular level, allele CoEEs are expected to involve biallelic expression, DAEEs could arise due to canonical genomic imprinting or other diverse allelic effects, and AAEEs are expected to involve random monoallelic effects for which maternal allele-expressing cells arise at the expense of paternal allele-expressing cells and vice versa (Figure 1A). F1 hybrid mice are derived from isogenic lines, eliminating allelic effects due to genetic variation. Further, by focusing on gene expression at the tissue level, transcriptional noise and bursting in individual cells is averaged across all cells in the tissue sample, eliminating this cellular noise (Piras and Selvarajoo, 2015).

For our screen, we initially focused on the dorsal raphe nucleus (DRN), which is the largest serotonergic nucleus innervating the forebrain and which regulates a variety of brain functions implicated in mental illness (Soiza-Reilly and Commons, 2014). We collected RNA-seq data for several F1cb (n = 9) and F1bc (n = 9) biological replicates for the adult female DRN and tallied maternal and paternal allele expression levels (Bonthuis et al., 2015). Females are used so that random monoallelic X-linked genes can be used as an internal control (see below). To analyze allele co-expression patterns across replicates, we center and whiten the allele expression data for the F1cb and F1bc crosses (see STAR Methods) and then compute the Pearson correlation for maternal and paternal allele expression levels across all 18 replicates for each gene (Figures 1B and 1C). We refer to the observed Pearson correlation between the two alleles as the "r_a" value (Figure 1C). Our approach revealed autosomal genes with putative allele CoEEs (*Fuca2*), DAEEs (*Bmp7*), or AAEEs (*Adora2b*) (Figure 1D). However, low r_a values indicating DAEEs or AAEEs could simply be caused by noise in the data. To address this problem, we devised an empirical Bayesian approach to estimate the 95% confidence interval (CI) for the ground-truth biological r_a value (r_{ab}) for any expressed gene (Figure 1E; also see STAR Methods and Figure S1). With our approach, we defined the r_{ab} 95% CI for each gene expressed in the adult female DRN and observed high-confidence AAEEs for random, monoallelic X-linked genes (Figure 1E; *Cd99l2*, *Mageh1*), DAEEs for known canonical imprinted genes (Figure 1E; *Zrsr1*, *Zdbf2*), and allele CoEEs for some autosomal genes (Figure 1E; *Hsd11b1*, *Vps4b*). Below, we use this general methodology to uncover diverse, non-genetic ASE effects in vivo in the mouse and primate.

Diverse Allelic Effects Exist In Vivo in the Mouse, and ASE Effects Predominate in the Neonatal Brain

We screened for non-genetic allelic effects in the mouse to uncover the diversity of allelic effects that exist in vivo, to discover the identity of the impacted genes, and to test whether ASEs are enriched in particular stages of brain development, brain regions, or tissue types. From the adult female DRN data, we defined genes with high-confidence AAEEs as those with an upper-bounded r_{ab} 95% CI $\leq$ 0, indicating a negative allele correlation (Figure 2A). To define criteria for high-confidence DAEEs and CoEEs, we examined the data for random monoallelic X-linked genes. We found that 94% of the 517 X-linked genes expressed in the adult female DRN have an upper-bounded 95% CI $\leq$ +0.75 (Figure 2B), indicating a threshold to detect random monoallelic effects across the different expression level

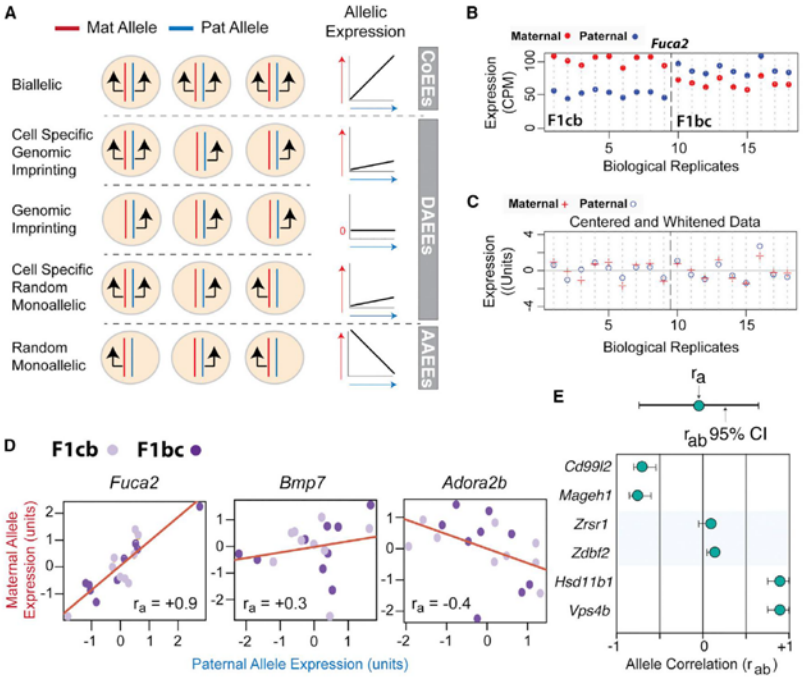


Figure 1. A Genomics and Statistical Approach to Screen for Diverse ASE Effects In Vivo

(A) Schematic of the expected relationship between different allelic effects at the cellular level and maternal and paternal allele-correlated expression patterns at the tissue level. Black arrows indicate the expressed alleles in individual cells. Linear regression line schematics depict the relative expression patterns of the two alleles across replicates (allelic expression), with the red arrow indicating increasing maternal allele expression and the blue arrow indicating increasing paternal allele expression. The slope of the regression line depicts the co-expression relationship between the two alleles. Biallelic expression promotes allele CoEEs, genomic imprinting promotes DAEEs, cell-type-specific genomic imprinting or random monoallelic expression promotes partial DAEEs, and random monoallelic expression in all cells promotes AAEEs (see text).

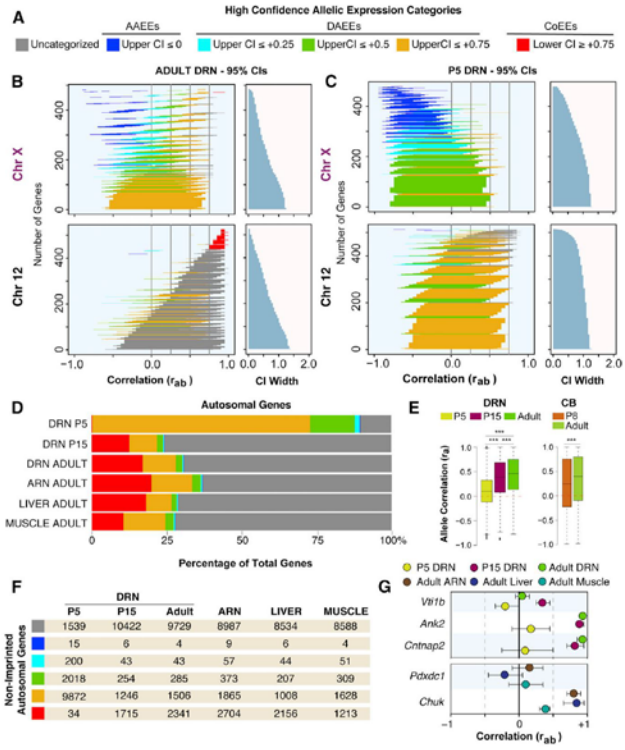
(B and C) Example of maternal (red dot) and paternal (blue dot) allele expression patterns across RNA-seq biological replicates (n = 18) for *Fuca2* in the DRN of adult female F1cb and F1bc mice (B). Note the correlated expression of the two alleles despite complex cross and parent-of-origin effects. After centering the F1cb and F1bc RNA-seq data, the variance in the expression levels of the maternal (red plus) and paternal (blue circle) alleles reveals correlated allelic expression patterns (r_a = +0.9) (C).

(D) RNA-seq analysis of maternal and paternal allele expression correlation (r_a) reveals genes with co-expressed alleles (*Fuca2*), genes with DAEEs (*Bmp7*), and genes with AAEEs (*Adora2b*).

(E) Schematic of the presentation of the observed r_a and estimated r_{ab} 95% CI for individual genes. Below are examples of random monoallelic X-linked genes with AAEEs (*cd99l2*, *Mageh1*), canonical imprinted genes with DAEEs (*Zrsr1*, *Zdbf2*), and autosomal genes with allele CoEEs (*Hsd11b1*, *Vps4b*).

and variance conditions in the data (see STAR Methods). Only one X-linked gene, *Ddx3x*, has a lower-bounded r_{ab} 95% CI $\geq$ +0.75, and this gene is known to escape X inactivation (Yang et al., 2010; Figure S4D). Therefore, we defined high-confidence allele CoEEs as genes with a lower-bounded r_{ab} 95% CI $\geq$ +0.75, indicating a robust positive correlation, and defined DAEEs as genes with an upper-bounded r_{ab} 95% CI $\leq$ +0.75, like X-linked genes (Figure 2A).

We further defined three subcategories of DAEEs. The first category includes genes with an upper-bounded 95% CI $\leq$ +0.75 (Figures 2A and 2B). The second and third categories capture more robust DAEEs for genes with an upper-bounded CI $\leq$ +0.5 and CI $\leq$ +0.25, respectively (Figures 2A and 2B). Notably, 100% of the canonical imprinted genes in the adult mouse DRN (Bonthuis et al., 2015) meet our criteria for high-confidence DAEEs or AAEEs. Further, six known monoallelic protocadherin genes (Hirano et al., 2012) have DAEEs, including *Pcdhβ12*, *Pcdhβ3*, *Pcdhβ15*, *Pcdhβ10*, *Pcdhβ6*, and *Pcdhβ20*, but none of the expressed biallelic *Pcdh* genes have DAEEs (Table S1). Genes with AAEEs or DAEEs are not enriched among genes with strong



	DRN					
Non-Imprinted Autosomal Genes	P5	P15	Adult	ARN	LIVER	MUSCLE
	1539	10422	9729	8987	8534	8588
	15	6	4	9	6	4
	200	43	43	57	44	51
	2018	254	285	373	207	309
	9872	1246	1506	1865	1008	1628
	34	1715	2341	2704	2156	1213

Figure 2. In Vivo Screen for ASE Effects in the Mouse Brain at Different Ages and Across Different Adult Tissues Reveals Genes with AAEEs, DAEEs, and Allele CoEEs and an Enrichment for DAEEs in the Neonatal Brain

(A) Criteria for categorizing genes according to AAEEs, DAEEs, and allele CoEEs; CI represents r_{ab} 95% confidence interval.

(B) Plots of the r_{ab} 95% CIs for all chromosome (chr)X and chr12 genes expressed in the adult female DRN reveal high-confidence AAEEs, subcategories of DAEEs, and high-confidence allele CoEEs. Gene CIs are colorized according to the appropriate category (A), and uncategorized genes are indicated in gray. The width of the CIs are presented in the right graph. Most X-linked genes have an upper-bounded 95% CI less than or equal to +0.75, indicating a threshold for ASE effects. Few autosomal genes achieve this strict threshold (see chr12 data).

(C) Plots of the r_{ab} 95% CIs for all chrX and chr12 genes expressed in the P5 female DRN reveal a substantial shift toward increased DAEEs among autosomal genes.

(D) The percentage of all expressed autosomal genes with different allelic effects in the P5, P15, and adult female DRN as well as in the adult female ARN, muscle, and liver. DAEEs are prevalent in the P5 DRN, and allele CoEEs are more prevalent in juvenile and adult tissues.

(E) Boxplots of the r_a values for all autosomal genes analyzed in the P5, P15, and adult DRN and in the P8 and adult cerebellum (CB). A significant genome-wide developmental change in ASE effects occurs, involving increased DAEEs at P5 and P8 and increased allele co-expression at P15 and in adults. DRN, one-way ANOVA, and Tukey's HSD post-test were used; CB, two-tailed t test. ***$p < 0.001$.

(F) Number of non-imprinted, autosomal genes in each allelic expression category (see A) for each age and tissue type.

(G) Examples of genes with DAEEs at all ages in the DRN (*Vti1b*) or age-specific DAEEs in the DRN (*Ank2*, *Cntnap2*); examples of genes that exhibit DAEEs across different tissue types (*Pdxdc1*) or in a tissue-specific manner (*Chuk*).

Cast or B6 allele expression biases (Figures S2A–S2C). This result supports the non-genetic nature of the allelic effects and is expected because Cast versus B6 allelic differences are the same in each of the F1 hybrid offspring and are not an expected source of DAEEs or AAEEs. Genes with AAEEs or DAEEs do not have reduced expression levels or fewer SNPs relative to other autosomal genes (Figures S2D–S2F); however, genes with allele CoEEs exhibit higher expression levels (Figure S2E). Overall, our screen has excellent sensitivity and specificity for the detection of diverse and robust non-genetic allelic effects in vivo.

In the adult female DRN, we found that 70% of the 13,909 autosomal genes analyzed do not meet criteria for high-confidence allelic effects (Figure 2B, e.g., chromosome [chr]12, gray CIs). Most autosomal genes have r_{ab} 95% CIs with an upper bound near +1.0 (Figure 2B, chr12, gray and red CIs), which reflects the expected trend toward allele CoEEs. Intriguingly however, our screen uncovered four autosomal genes with high-confidence AAEEs (0.02%), 1,834 genes with DAEEs (13%), and 2,341 genes (16%) with allele CoEEs. Therefore, we found 4,179 autosomal genes that are distinguished by their unique allelic effects in the adult DRN.

To compare ASEs at different stages of brain development, we performed our screen in the DRN of post-natal day (P)5 neonates and P15 juveniles and contrasted the results to adults. Remarkably, 88% of autosomal genes in the P5 DRN exhibit high-confidence DAEEs (Figure 2C) compared to only 11% in the P15 DRN and 13% in the adult DRN (Figure 2B), revealing a profound increase in DAEEs in neonates (Figure 2D). The number of genes with high-confidence AAEEs did not change substantially between P5 (0.2%), P15 (0.1%), and adult (0.08%) offspring (Figures 2D and 2F). However, the number of genes with allele CoEEs increased by over ~50-fold in P15 juveniles and adults compared to P5 neonates (Figures 2D and 2F). For example, *Vti1b* exhibits DAEEs in the P5, P15, and adult DRN, while

Ank2 and *Cntnap2* exhibit DAEEs at P5 but allele CoEEs at P15 and in adults (Figure 2G). To independently validate this surprising developmental change, we analyzed independently generated F1cb and F1bc RNA-seq datasets for the female P8 and P60 cerebellum (Perez et al., 2015). Boxplots of the observed r_a values for autosomal genes confirmed a statistically significant age effect involving a genome-wide shift from lower r_a values in neonates to higher r_a values in older mice in both the DRN and cerebellum (Figure 2E). Therefore, independent studies of different brain regions show a developmental shift from DAEEs in the neonatal brain to increased allele CoEEs in the adult brain.

Genomic imprinting effects are enriched in the brain (Babak et al., 2015; Bonthuis et al., 2015; Perez et al., 2015), and therefore we tested whether the brain is generally enriched for non-genetic ASE effects. We compared allelic effects in the adult female DRN and arcuate nucleus (ARN) of the hypothalamus to liver (endoderm-derived) and skeletal muscle (mesoderm-derived). The results uncovered autosomal genes with high-confidence AAEEs, DAEEs, and allele CoEEs in each brain region (DRN and ARN) and tissue type (brain, liver, and muscle) (Figures 2D and 2F). Very few autosomal genes exhibit AAEEs in any tissue (or age) (Figure 2F). We found that ~17% of autosomal genes in the ARN and muscle exhibit DAEEs compared to 11% in liver and 7% in the DRN, indicating modest brain region and tissue differences. Genes with high-confidence DAEEs in all tissues, such as *Pdxdc1*, were identified, and so were genes with putative tissue differences in allelic effects, such as *Chuk*, which has DAEEs in the liver but not in the muscle or ARN (Figure 2G). Overall, an enrichment for non-genetic ASE effects in the brain was not observed, and diverse allelic effects are present in all major tissue types in the mouse (Table S1).

We sought to define the subset of autosomal genes with ASE effects not due to imprinting. Previously, we found that less than ~2% of genes exhibit imprinting in the adult female DRN, ARN, liver, and muscle (Bonthuis et al., 2015). Here, we also profiled imprinting effects in the P5 and P15 female DRN (Table S1). With these imprinting datasets, we determined that only ~6.9% and ~1.5% of all autosomal genes with AAEEs and DAEEs, respectively, are imprinted (Figure S3A and Table S1). Therefore, most of the ASE effects detected in our screen are unrelated to imprinting. We found that these allelic effects are highly reproducible by pyrosequencing (see supplemental data and Figures S3B–S3D). Finally, we unexpectedly discovered developmental changes to the allelic expression of X-linked genes in the female brain; these results are also presented in a supplemental data section (supplemental data and Figure S4). Below, we investigate the cellular nature of autosomal allele CoEEs, DAEEs, and AAEEs in the brain.

Allele CoEEs, DAEEs, and AAEEs Involve Differences in the Prevalence of Biallelic versus Monoallelic Brain Cells In Vivo

We hypothesized that DAEEs are related to random monoallelic expression in subpopulations of cells (Figure 3A). To discover the cellular nature of AAEEs, DAEEs and allele CoEEs, we resolved allelic expression at the cellular level in mouse brain tissue sections by designing RNAscope probes to rapidly processed introns in the nascent RNA of target genes (Bonthuis et al., 2015). With this approach, we quantified putative monoallelic and biallelic cells in the female DRN and ventral periaqueductal gray region in isogenic B6 adult females for ten randomly selected genes with r_a values ranging from +0.99 to −0.4. Our results reveal that genes with DAEEs or AAEEs exhibit monoallelic expression in more brain cells than genes with allele CoEEs (Figure 3B). Indeed, a linear regression analysis found that a lower r_a value is associated with a significant increase in the percentage of monoallelic brain cells (Figure 3B). For example, *Bmp4* has a low r_a value of −0.33 and exhibits monoallelic expression in 76% of positive cells and biallelic expression in 24% of positive cells (Figures 3B and 3C). In contrast, *Mtap1b* has an r_a value of +0.99 and exhibits biallelic expression in 66% of positive cells (Figures 3B and 3D).

We previously found that ~25% of apparent monoallelic cells are potentially due to cryosectioning artifacts (Bonthuis et al., 2015). To determine whether the monoallelic cells observed for genes with high r_a values are artifacts, we co-labeled with the biallelic internal control gene and neuron marker *Syn2* (Bonthuis et al., 2015). Co-labeling for *Mtap1b*+*Syn2* in the DRN revealed subpopulations of neurons in the midbrain that exhibit biallelic *Syn2* expression but monoallelic expression for *Mtap1b* (Figure 3E), suggesting that not all of the monoallelic cells for these genes are artifacts. Indeed, some genes with high r_a values, such as *Atp8a1*, have relatively high numbers of apparent monoallelic cells (Figure 3B). We expect that these apparent monoallelic effects are due to transcriptional bursting. Bursting effects are averaged across cells in the tissue in our RNA-seq data, revealing the allele CoEEs at the mRNA and tissue level.

Nr1d1 and *Cacng2* are two autism-linked genes that exhibit DAEEs in the P5 DRN but increased allele co-expression in the P15 and adult brain (Figure 3F). For these genes, we found a significant decrease in the number of monoallelic cells and an increase in the number of biallelic cells in adults compared to P5 offspring (Figures 3G and 3H). As a control, *Oxtr* exhibits similar DAEEs at all ages tested (Figure 3F) and no significant change in the prevalence of monoallelic versus biallelic cells in the P5 versus adult DRN (Figures 3G and 3H). Overall, we conclude that AAEEs and DAEEs involve more monoallelic expression at the cellular level, while allele CoEEs involve more biallelic expression.

DAEEs Shape Genetic Architecture at the Cellular Level In Vivo

We next tested the hypothesis that AAEEs and DAEEs cause mosaics of monoallelic mutant (monoallelic-MT) and monoallelic wild-type (monoallelic-WT) allele-expressing cells for heterozygous mutations in the brain. We obtained heterozygous knockout-*LacZ* reporter mice on a B6 background for three genes with DAEEs in the P5 and adult brain, including *Bmp7*, *Bmp4*, and *Adora2b* (Figure 3B and Table S1). To detect the cellular expression of WT versus MT alleles in each heterozygous line, we designed RNAscope in situ hybridization probes to detect mRNA from the WT alleles and a probe to detect the *LacZ* mRNA from the MT allele. RNAscope probes are semi-quantitative, and each staining dot corresponds to a single transcript. By co-labeling with probes targeting the WT and MT

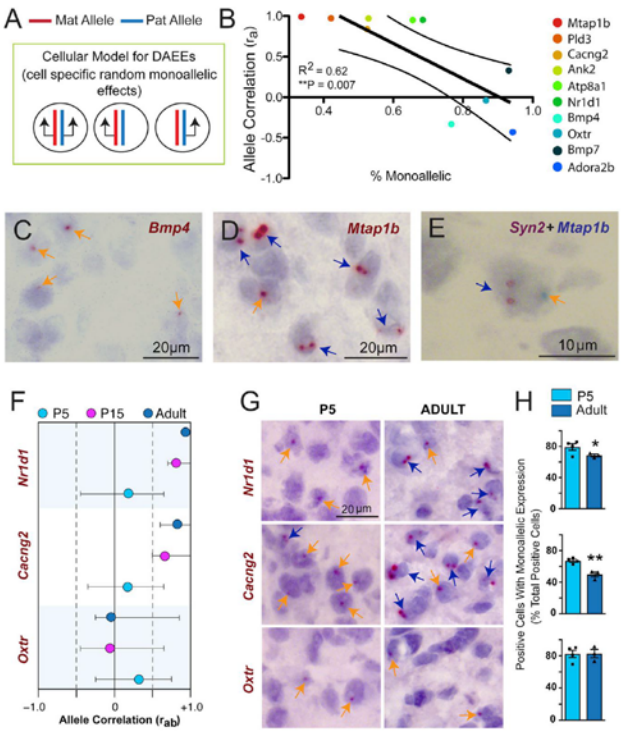


Figure 3. Genes with DAEEs Exhibit Increased Monoallelic Expression at the Cellular Level in the Brain Compared to Genes with Co-expressed Alleles

(A) Schematic hypothesis indicates that non-imprinted, autosomal genes with DAEEs may exhibit random monoallelic effects in some cells and biallelic expression in others in the brain.

(B) For ten randomly selected, non-imprinted autosomal genes with different r_a values, allele-specific nascent RNA in situ hybridization analysis was performed in tissue sections of the DRN and in the ventral periaqueductal gray region of the midbrain from adult female B6 mice. The percentage of monoallelic cells from the total number of positive cells (x axis) is plotted as a function of the r_a value (y axis) for each gene. Linear regression reveals a significant negative correlation, indicating that genes with lower r_a values are associated with more monoallelic brain cells. Approximately 30% of monoallelic cells are false due to partial nuclei from cryosectioning, and therefore the x axis begins at 30%.

(C and D) Examples of allelic, nascent RNA in situ hybridization staining in the adult female B6 DRN and ventral periaqueductal gray region. Genes with low r_a values, such as *Bmp4* (C), predominantly exhibit monoallelic expression at the cellular level (orange arrows), while genes with high r_a values, such as *Mtap1b* (D), predominantly exhibit biallelic expression (blue arrows).

(E) Co-labeling of *Mtap1b* (blue staining) with the biallelic neuron marker gene *Syn2* (brown staining) reveals subpopulations of neurons with putative monoallelic *Mtap1b* and biallelic *Syn2* expression.

(F–H) Plots of the r_a values and r_{ab} 95% confidence intervals for two genes that shift from DAEEs at P5 to allele co-expression in the adult DRN (*Nr1d1* and *Cacng2*) and one gene that exhibits DAEEs at P5 and P15 and a trend toward DAEEs in adults (*Oxtr*) (F). Examples of monoallelic and biallelic cells in the P5 versus adult female B6 DRN and ventral periaqueductal gray region (G). More monoallelic cells are observed in the P5 than the adult for *Nr1d1* and *Cacng2*, but not for *Oxtr*. Cell counts reveal significantly more monoallelic than biallelic cells in P5 neonates compared to adults for *Nr1d1* and *Cacng2*, but no difference for *Oxtr*, consistent with r_{ab} data for these genes (H). N = 4, Student's t test, *$p < 0.05$, **$p < 0.01$.

alleles in the heterozygous mice, we compared the relative expression of the two alleles at the cellular level in the brain.

We first analyzed midbrain tissue sections in P5 and adult *Bmp7*$^{+/LacZ}$ mice by co-labeling with probes targeting *Bmp7* and *LacZ* mRNA, and we uncovered subpopulations of cells that preferentially express the MT (*LacZ*-expressing) allele (Figures 4A' and 4B'), both alleles (Figures 4A'' and 4B''), or the WT allele ((Figures 4A''' and 4B'''). The probe targeting *LacZ* mRNA yielded very little non-specific staining in *Bmp7*$^{+/+}$ brain tissue (Figure S5). Further, the probe targeting WT *Bmp7* mRNA did not exhibit non-specific staining in *Bmp7*$^{LacZ/LacZ}$ homozygous brain tissue (Figure S5) and robustly labeled the choroid plexus in E14.5 *Bmp7*$^{+/+}$ embryos (Figure S5), which highly express *Bmp7* (Segklia et al., 2012). Thus, our approach accurately detects expression of MT and WT alleles and reveals that DAEEs can shape genetic architecture at the cellular level in the brain. A similar analysis of MT and WT allelic expression in *Bmp4*$^{+/LacZ}$ (Figures 4C and 4D) and *Adora2b*$^{+/LacZ}$ heterozygous mice (Figures 4E and 4F) also uncovered brain cells that preferentially express either the MT or WT allele as well as biallelic cells, indicating that the result generalizes to other genes with DAEEs.

In our analysis, we observed five general cell classes, including biallelic cells, cells with a MT or WT allele bias, and monoallelic cells that only express the MT or the WT allele (Figure 5A). We counted the relative numbers of these five cell classes in the adult female DRN/PAG, the arcuate nucleus (ARN) region of the hypothalamus, the cortex (CTX), and in the meninges. In the *Bmp7*$^{+/LacZ}$, *Bmp4*$^{+/LacZ}$, and *Adora2b*$^{+/LacZ}$ mice, we found that positive cells in the DRN/PAG, ARN/HYP, and CTX are predominantly monoallelic for either the WT or MT allele; a subpopulation of cells exhibit allele expression biases; and a smaller subpopulation exhibit biallelic expression (Figure 5B). We further found that the number of WT- versus

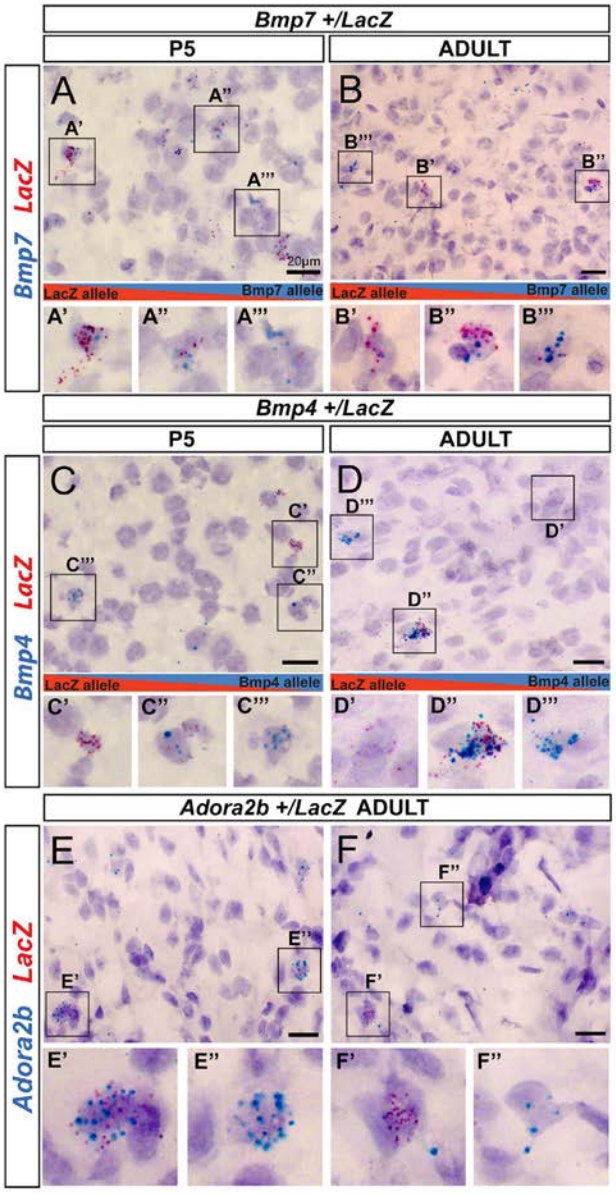
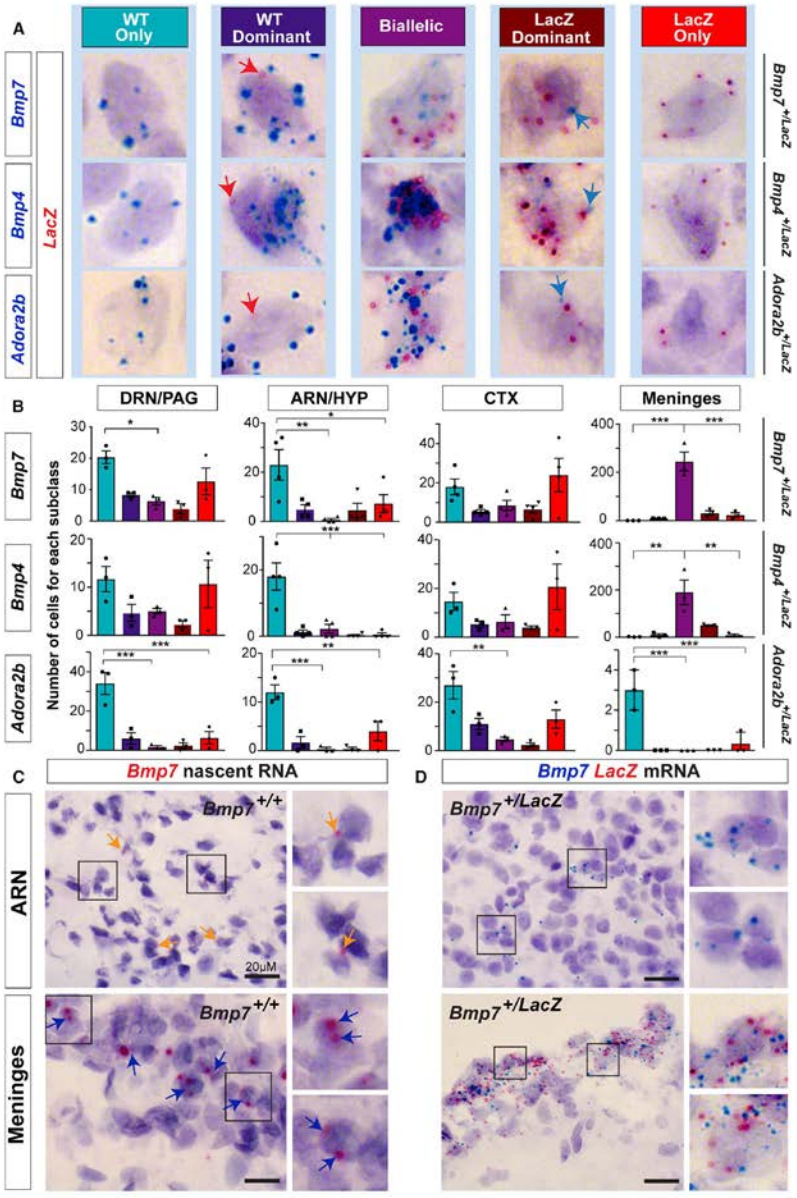


Figure 4. DAEEs Cause Mosaics of Brain Cells that Differentially Express Mutant versus Wild-Type Alleles for Inherited Heterozygous Mutations

(A and B) An analysis of the cellular expression of mutant (MT) versus wild-type (WT) alleles in *Bmp7*$^{+/LacZ}$ knockout-reporter heterozygous mice at P5 (A) and in adults (B) in the DRN and periaqueductal gray region of the midbrain. Semiquantitative RNAscope in situ hybridization probes targeting the mRNA from the MT *LacZ* allele (red) versus the WT *Bmp7* allele (blue) reveal subpopulations of brain cells that preferentially express the MT allele (A' and B') and the WT allele (A''' and B''') and biallelic cells that express both alleles (A'' and B'').

(C–F) The cellular expression of MT versus WT alleles in *Bmp4*$^{+/LacZ}$ knockout-reporter heterozygous mice at P5 (C) and in adults (D) and in adult *Adora2b*$^{+/LacZ}$ mice (E and F). Subpopulations of cells that preferentially express the MT allele (C', D', and F') are present in the brain for each mouse line, as are cells that preferentially express the WT allele (C''', D''', E'', and F'') and biallelic cells (C'', D'', and E').



(legend on next page)

CellPress

LacZ-allele-expressing cells in the DRN/PAG and CTX is not significantly different for *Bmp7* and *Bmp4* (Figure 5B); however, a significant bias to monoallelically express the WT allele was observed in the ARN/HYP for both genes (Figure 5B), revealing a difference in the genetic architecture between the brain regions. For *Adora2b*, a significant bias to express the WT allele was observed in the DRN/PAG, ARN/HYP, and meninges, and a trend exists in the CTX (Figure 5B). Thus, the relative prevalence of monoallelic-MT and monoallelic-WT brain cells can be gene, cell, and brain region specific.

In contrast to the cells analyzed in other brain regions, the majority of *Bmp7*+ and *Bmp4*+ meningeal cells are biallelic (Figure 5B). We investigated these apparent cell-type-specific allelic effects with two independent approaches. First, *Bmp7*+ cells in the ARN region exhibit monoallelic expression by nascent RNA in situ hybridization in $Bmp7^{+/+}$ mice (Figure 5C) and by mRNA in situ hybridization in $Bmp7^{+/LacZ}$ mice (Figure 5D). In contrast, in meningeal cells, nascent RNA in situ revealed predominantly biallelic *Bmp7* expression (Figure 5C), and most meningeal cells expressed both the WT *Bmp7* and MT *LacZ* mRNA in $Bmp7^{+/LacZ}$ mice (Figure 5D). Therefore, two independent methods indicate that DAEEs can be regulated in a cell-type-specific manner in the brain.

Notably, major changes to the proportion of monoallelic-MT and monoallelic-WT cells in reciprocal heterozygotes were not observed, ruling out imprinting as a plausible mechanism (Figures S6A–S6C). In addition, we tested whether ectopic expression of the *LacZ* transgene could be occurring by comparing the numbers of WT-allele-expressing cells in B6 control mice to the total numbers of monoallelic-MT, monoallelic-WT, and biallelic cells in the heterozygotes. No significant difference was observed (Figures S6E and S6F), indicating that the *LacZ* allele is not ectopically labeling an additional cell population. Overall, DAEEs can shape genetic architecture at the cellular level in the brain.

Non-genetic DAEEs Exist in the Primate Brain

Our screen uncovered diverse non-genetic allelic effects in the mouse brain, but we do not know whether similar effects exist in the primate brain. Therefore, we analyzed allele co-expression in the DRN region of cynomolgus macaques (*Macaca fascicularis*), which are a well-established primate model for humans with a highly polymorphic genome (Higashino et al., 2012; Yan et al., 2011). For ten mother-father-daughter triads, we identified polymorphic sites that can discriminate maternal and paternal alleles in the daughters by performing whole-genome sequencing on DNA extracted from blood samples from the parents for each triad (Figure 6A). Next, to measure allelic expression in the daughters, we performed whole-transcriptome profiling on DRN RNA extracted from juveniles (3 to 20 months of age) (Figure 6A and STAR Methods).

We are able to analyze allele co-expression for 821 autosomal genes and 17 X-linked genes because they have allele-discriminating SNPs in all ten daughters (Table S4). For these 838 genes, we calculated the r_a statistic and uncovered autosomal genes with putative allele CoEEs, such as *AGAP1*, or DAEEs, such as *CNTN1* (Figure 6B). However, unlike inbred mice, apparent DAEEs could arise in primates due to genetic variation in regulatory elements. Therefore, we modified our statistical framework to estimate the impact of genetic variation, biological variation, and technical noise in our primate data, and we defined the r_{ab} 95% CI for the ground-truth, non-genetic allelic effects exhibited by each gene (see STAR Methods).

We found that the r_{ab} 95% CIs for X-linked genes have reduced values compared to autosomal genes; this is consistent with random monoallelic effects (Figure 6C). Most X-linked genes (88%) have an upper-bounded r_{ab} 95% CI less than or equal to +0.75, indicating a threshold for high-confidence primate DAEEs. We found that two autosomal genes (0.2% of autosomal genes tested) meet this threshold: *RBM48* (RNA binding motif protein 48) and *HTT* (Huntingtin) (Figures 6C and 6D). *RBM48* is a protein-coding gene with largely unknown functions, while *HTT* is mutated in Huntington's disease (HD). HD is an autosomal-dominant disorder, and epigenetic allelic effects for *HTT* could potentially have important implications. We did not identify significant imprinting effects for any of the 838 genes (5% FDR threshold), indicating that the DAEEs are not due to genomic imprinting. Therefore, DAEEs are a conserved phenomenon between mice and primates. In addition, we identified 213 autosomal genes with high-confidence allele CoEEs (Figure 6C). By considering our results as a random sampling of the 13,030 annotated autosomal genes expressed in the juvenile macaque DRN, we estimate a genome-wide total of 26 and 3,387 genes with high-confidence DAEEs and CoEEs, respectively.

In our analysis of the primate data, we identified 68 genes with an upper-bounded r_{ab} 95% CI $\leq$ +0.95 (Figure 6C, yellow CIs). These genes do not meet the criteria for high-confidence DAEEs; however, since 92% of autosomal genes have an upper-bounded CI equal to +1 (Figure 6C), an upper bound of less than +0.95 suggests at least partial non-genetic DAEEs. In support of this interpretation, we observed that two X-linked genes also have an upper bound of less than +0.95 (Figure 6C). We call these putative DAEEs to distinguish them from the more robust DAEEs detailed above. For example, *NARS* and *TPNC1* exhibit putative DAEEs, which are clearly distinct from both the

Figure 5. DAEEs Involve Brain-Region-Specific and Cell-Type-Specific Allele-Silencing Effects that Shape Genetic Architecture in the Brain

(A) RNAscope in situ hybridization labeling for *Bmp7*, *Bmp4*, or *Adora2b* (blue staining) and *LacZ* (red staining) mRNA in the brain reveals subclasses of positive cells according to the relative allelic expression of the WT or MT *LacZ* allele in $Bmp7^{+/LacZ}$, $Bmp4^{+/LacZ}$, and $Adora2b^{+/LacZ}$ adult female mice.

(B) Numbers of monoallelic (WT or MT *LacZ* allele), dominant-allele biased (WT or MT *LacZ* allele), and biallelic cells in the DRN-PAG, ARN-HYP, CTX, and meninges in $Bmp7^{+/LacZ}$, $Bmp4^{+/LacZ}$, and $Adora2b^{+/LacZ}$ knockin-knockout adult mice. One-way ANOVA with Tukey's HSD post-test was used; $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

(C and D) Monoallelic (orange arrows) and biallelic (blue arrows) *Bmp7*+ cells identified by nascent RNA in situ in the ARN-HYP and in meningial cells in B6 adult female $Bmp7^{+/+}$ mice (C). Monoallelic and biallelic cells are also detected in the ARN/HYP and meninges, respectively, by mRNA in situ hybridization for *Bmp7* (blue) and *LacZ* (red) in adult $Bmp7^{+/LacZ}$ mice (D).

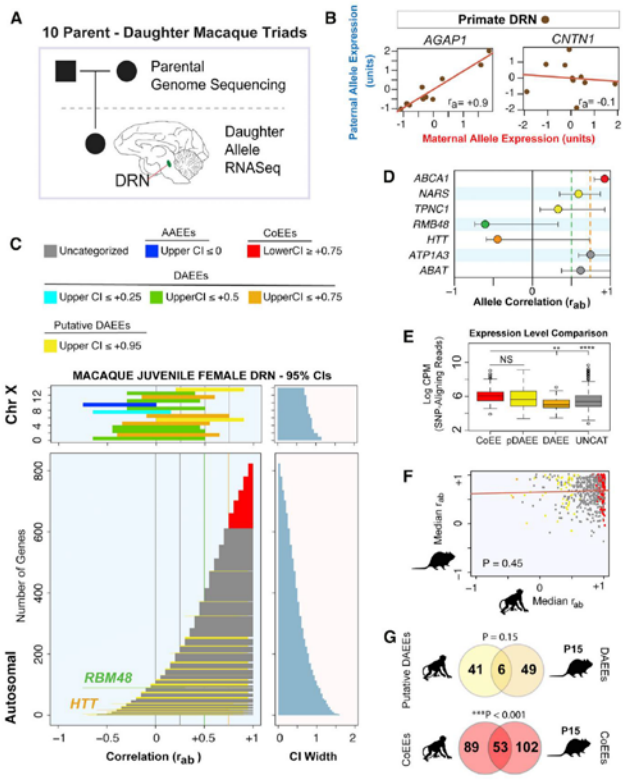


Figure 6. Identification of Autosomal DAEEs and Allele CoEEs in the Primate Brain

(A) Schematic of the strategy to profile allele co-expression in the DRN region of juvenile female cynomolgus macaques. For ten parent-offspring trios, we performed whole-genome sequencing of the parents and transcriptome sequencing of RNA extracted from the DRN of the daughters. SNPs that distinguish maternal from paternal alleles in the daughters are determined from the parental genomes and RNA-seq datasets.

(B) By analyzing r_a values for the 838 genes with maternal and paternal allele RNA-seq data in all ten daughters, we defined genes with allele CoEEs (*AGAP1*) and potential DAEEs (*CNTN1*) in the juvenile female macaque DRN.

(C) Plots of the non-genetic r_{ab} 95% CIs for 17 X-linked genes and 821 autosomal genes, colorized according to the categories of high-confidence allelic effects indicated in the legend. Most X-linked genes exhibit AAEEs or DAEEs. Most autosomal genes exhibit allele CoEEs (red) or do not exhibit sufficiently robust allelic effects to be categorized with high confidence (gray); however, high-confidence DAEEs were discovered for *RBM48* and *HTT*. In addition, more modest putative DAEEs were observed for several autosomal genes (see main text).

(D) The r_a values and non-genetic r_{ab} 95% CIs for examples of primate genes with allele CoEEs (*ABCA1*), putative DAEEs (*NARS*, *TPNC1*), high-confidence DAEEs (*RBM48*, *HTT*), and uncategorized genes (*ATP1A3*, *ABAT*).

(E) Boxplots comparing the primate DRN expression level of autosomal and X-linked genes with allele CoEEs, putative DAEEs (pDAEEs), DAEEs, and uncategorized (UNCAT) genes. A significant main effect of gene class was observed (one way ANOVA, $p < 0.0001$), and a Tukey's HSD post-test revealed no difference between genes with CoEEs versus putative DAEEs, but genes with CoEEs are expressed at a higher level than genes with DAEEs or uncategorized genes. $^{**}p < 0.01$, $^{***}p < 0.001$. Expression level is based on SNP-aligning reads only. CPM represents counts per million reads.

(F) A comparison of the median r_{ab} values extracted from the r_{ab} 95% CIs for autosomal genes examined for allelic effects in the macaque compared to the values in the mouse for the orthologs. A Spearman rank partial-correlation analysis controlling for expression levels reveals no significant conservation for allelic effects at the gene level between the two species ($p = 0.45$).

(G) Venn diagram indicates that six genes with putative DAEEs in the primate also have DAEEs in the P15 mouse DRN, while 53 genes with allele CoEEs in the primate also have CoEEs in the mouse. The overlap for CoEEs between the species is greater than expected by chance, indicating some conservation of these effects (hypergeometric test p value shown).

allele CoEEs exhibited by *ABCA1* and the uncategorized allelic effects for *ATP1A3* and *ABAT* that involve an upper r_{ab} CI equal to 1.0+, like most genes (Figure 6D). Genes with putative DAEEs do not exhibit reduced expression levels compared to genes with allele CoEEs or uncategorized genes (Figure 6E), indicating that their allelic effects are a distinguishing feature. We estimate a genome-wide total of 1,042 autosomal genes (8%) with putative DAEEs in the primate DRN, and we investigate these in detail below.

To test whether allelic effects are conserved at the gene level between primates and mice, we identified non-imprinted, autosomal genes with allelic expression data and orthologs in both species. We derived the median r_{ab} value for each gene from the P15 mouse DRN and juvenile primate DRN r_{ab} 95% CI modeling data. A partial correlation analysis comparing these median r_{ab} values between the two species after controlling for gene expression levels did not uncover a statistically significant correlation (Figure 6F). Indeed, neither *RBM48* nor *HTT* exhibit DAEEs in the mouse, and only six genes with putative DAEEs in the primate exhibit DAEEs in the mouse (Figure 6G). We did find a statistically significant overlap between the two species for the genes with allele CoEEs (Figure 6G). In summary, DAEEs exist in the primate brain, but largely impact different genes than in the mouse.

Putative DAEEs Impact Genes Linked to Mental Illness in the Macaque and Human Brain

Having determined that DAEEs exist in the primate DRN, we set out to test whether genes that are risk factors for mental illness are impacted. Using the DisGeNet Database (Bauer-Mehren et al., 2010), we identified risk genes for six different major forms of mental illness, including autistic disorders (ASD), schizophrenia (SCZ), bipolar disorder (BP), unipolar depression (UD), intellectual disability (ID), and attention deficit hyperactivity disorder (ADHD). Among the 821 autosomal genes analyzed in the primate DRN, we found high-confidence DAEEs or putative DAEEs for 7% of the 33 ASD risk genes analyzed, 4% of the 51 BP risk genes, 15% of the 33 ID risk genes, 10% of the 103 SCZ risk genes, 26% of the 15 UD risk genes, and 13% of the 23 ADHD risk genes (Figure 7A). We also identified subsets of risk genes with allele CoEEs (Figure 7A). Overall, putative DAEEs impact some mental illness risk genes in the primate brain.

We sought to test whether putative DAEEs are associated with an increased prevalence of monoallelic brain cells and simultaneously to determine whether the results from our macaque study are relevant to the human brain. We focused on two genes, *DEAF1* and *CNTNAP2*. *DEAF1* exhibits putative DAEEs in the macaque DRN (Figure 7B) and is linked to autism, language impairment, and intellectual disability in humans (Rajab et al., 2015; Vulto-van Silfhout et al., 2014). *CNTNAP2* exhibits a trend toward allele CoEEs in the macaque DRN (Figure 7B), exhibits high-confidence allele CoEEs in the mouse DRN (Figure 2G), and is also linked to autism, language impairment, and intellectual disability in humans (Alarcón et al., 2008; Arking et al., 2008). Using nascent RNA in situ hybridization, we contrasted the allelic expression of these two autism genes at the cellular level in post-mortem human brain tissue for four brain regions from control adult females aged 49 to 68 years. The brain regions we studied are implicated in ASDs (Amaral et al., 2008) and include the DRN (n = 5), orbitofrontal cortex (OFC) (n = 5), anterior cingulate cortex (ACC) (n = 12), and Broca's Area (BA) (n = 5) (Figure 7C).

We first tested whether allelic effects are consistent or variable between different individuals by counting the number of monoallelic versus biallelic *CNTNAP2*+ and *DEAF1*+ cells in 1 cm^2 sections of the ACC for twelve unrelated females. The results reveal that the total number of positive cells per cm^2 is highly variable between individuals, but hundreds to thousands of positive cells could be analyzed for allelic expression in each individual, with the exception of four brains that did not have *DEAF1*+ cells in the ACC (Figure 7D). Positive controls confirmed that the RNA was intact for each sample (not shown). We found that ~80% of *DEAF1*+ cells in the ACC are monoallelic, while only ~13%–30% of *CNTNAP2*+ cells are monoallelic, and the majority are biallelic in each individual (Figures 7C and 7D). Therefore, for at least some cases, genes with putative DAEEs are associated with more monoallelic brain cells than genes that do not exhibit these effects, and the effects are reproducible between individuals, indicating that genetic variation in regulatory elements is not causing the effects.

We extended our analysis to include different brain regions, including the DRN, OFC, and BA, and found that *DEAF1* exhibits monoallelic expression in ~80% of cells in each brain region (Figure 7E). However, for *CNTNAP2*, we found a statistically significant main effect of brain region ($p < 0.0001$, one-way ANOVA). We learned that the proportion of monoallelic *CNTNAP2*+ cells is greater in the OFC than in the ACC and BA and is greater in the DRN than in the ACC (Figure 7E). By performing co-labeling between *CNTNAP2* and the biallelic neuron marker gene *SYN2*, we found neuronal nuclei with biallelic *SYN2* expression and monoallelic *CNTNAP2* expression (Figure 7F), indicating potential monoallelic *CNTNAP2*+ neurons. In addition, we identified a small subpopulation of biallelic *DEAF1*+ neurons (Figure 7F). Overall, our results uncover non-genetic allelic effects in the primate brain that can impact genes linked to mental illness and differ between cells and brain regions.

DISCUSSION

Epigenetic ASE effects, such as random X inactivation and genomic imprinting, can profoundly influence the genetic architecture of human disease and brain disorders, though the spectrum of non-genetic ASE effects that exist in vivo has been unclear. Here, we devised a genomics and statistical framework to perform in vivo screening for diverse non-genetic allelic effects in the mouse and primate brain. By screening for allelic effects at different developmental stages and in different brain regions and tissue types in the mouse, we uncovered hundreds of autosomal genes with high-confidence DAEEs or AAEEs that are not due to genomic imprinting or genetic variation and are reproducible by pyrosequencing. Remarkably, DAEEs are revealed to be highly prevalent in the neonatal brain, while allele CoEEs are more prevalent at later developmental stages. At the cellular level, genes with DAEEs or AAEEs are associated with more monoallelic brain cells, while genes with allele CoEEs are associated with more biallelic brain cells. We determined that DAEEs interact with inherited heterozygous variants to cause mosaics of monoallelic-MT, monoallelic-WT, and biallelic brain cells in vivo. These effects can be cell type specific, since meningeal cells exhibit biallelic expression for *Bmp4* and *Bmp7*, while other brain cell populations are predominantly monoallelic. Finally, from a genomics study of macaque parent-offspring triads and a histological study of human brain tissue sections, we provide evidence that non-genetic autosomal DAEEs that are not due to imprinting exist in the primate brain, impact genes linked to mental illness, and are associated with reproducible, monoallelic expression effects at the cellular level. Our findings have potential implications for brain genetics, as discussed below.

Potential Mechanisms Regulating AAEEs, DAEEs, and Allele CoEEs

The mechanisms underlying most of the different allelic effects we uncovered are currently unclear. Moreover, the relationship between the AAEEs and DAEEs defined in vivo in our study and the autosomal random monoallelic effects that have been described in cell lines is also unclear (Eckersley-Maslin et al., 2014; Gendrel et al., 2014; Gimelbrant et al., 2007; Jeffries et al., 2013; Zwemer et al., 2012). Some random monoallelic genes identified in mouse neuronal progenitor cell lines

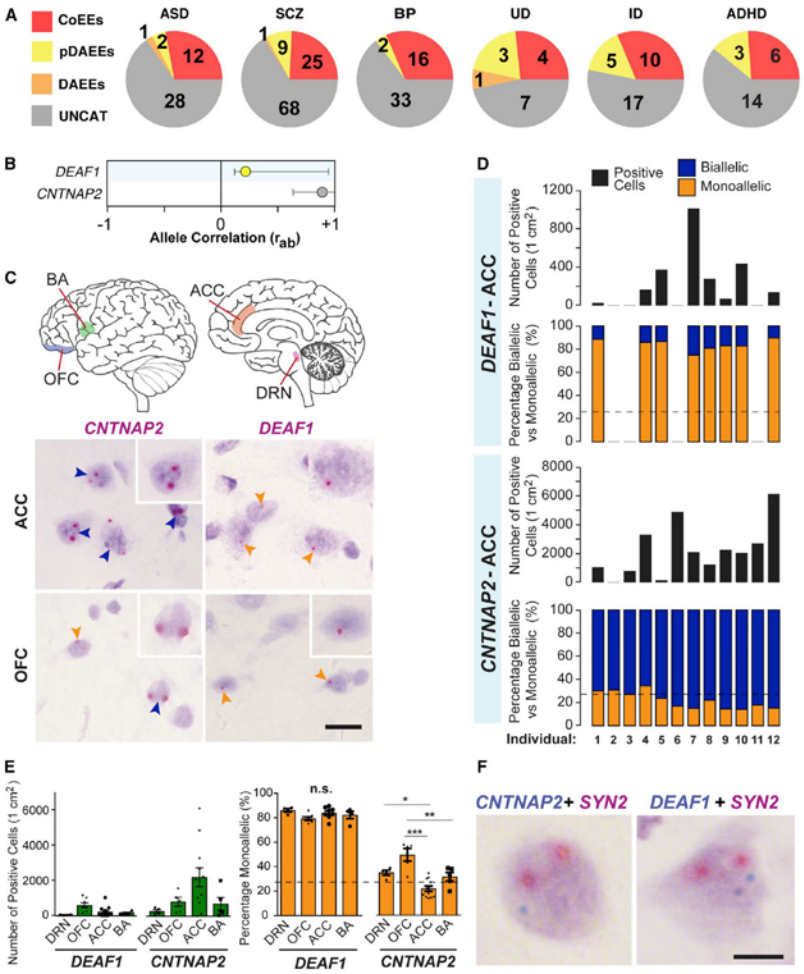


Figure 7. DAEEs Impact Genes Linked to Mental Illness in the Macaque and Human Brain

(A) Pie charts of the numbers of mental illness genes found with high-confidence CoEEs, DAEEs, or putative DAEEs in the macaque DRN. Shown are results for genes linked to autism spectrum disorder (ASD), schizophrenia (SCZ), bipolar disorder (BP), intellectual disability (ID), unipolar depression (UD), and attention deficit hyperactivity disorder (ADHD).

(B) The primate DRN r_a values and non-genetic r_{ab} 95% CIs for the autosomal autism genes *CNTNAP2* (uncategorized) and *DEAF1* (putative DAEEs).

(C) Nascent RNA in situ hybridization in four human brain regions was performed for *CNTNAP2* (allele CoEEs in macaque and mouse) and *DEAF1* (DAEEs in macaque) to determine the proportion of biallelic versus monoallelic cells. Examples of allele expression staining for each gene are shown for the ACC and OFC. *CNTNAP2* exhibits biallelic expression (dark blue) in most cells in the human ACC, while *DEAF1* exhibits monoallelic expression (orange arrows). In the OFC, *CNTNAP2* is predominantly biallelic, but monoallelic cells are observed, and *DEAF1* is predominantly monoallelic. OFC represents orbitofrontal CTX; BA represents Broca's Area; ACC represents anterior cingulate cortex.

(D) Analysis of the number of *CNTNAP2*+ and *DEAF1*+ cells in 1 cm^2 cryosections of the ACC for 12 adult control females and the proportion of biallelic versus monoallelic cells. The results reveal that the total number of *Cntnap2*+ and *Deaf1*+ cells per cm^2 is highly variable between individuals, but the relative proportion

(legend continued on next page)

(Gendrel et al., 2014) have significantly reduced r_a values in our mouse brain data and may exhibit ASE effects in vivo in the post-natal mouse DRN, but many do not (Table S3). More striking is that we show DAEEs to be highly prevalent in the developing neonatal brain, while allele CoEEs are more prevalent at later developmental stages, and that random monoallelic effects in embryonic stem cell lines were previously shown to increase in frequency following differentiation into neural progenitors (Eckersley-Maslin et al., 2014; Gendrel et al., 2014). Thus, our in vivo findings and previous in vitro studies both indicate that developmental gene expression programs involve complex epigenetic, allelic effects that might shape the impact of genetic variants on cell differentiation and survival during brain development.

We found that very few autosomal genes exhibit high-confidence AAEEs in vivo; however, AAEEs are relatively prevalent for X-linked genes, as expected for random X inactivation. Our data, therefore, indicate that autosomal, random monoallelic effects that are similar to X inactivation are very rare in vivo. These findings are potentially consistent with a recent study using single-cell RNA-seq; it concluded that the clonal random monoallelic effects observed in some cell lines are rare in primary fibroblasts and T cells (Reinius et al., 2016). In the future, in vivo lineage-tracing studies could more directly investigate the clonal relationships between cells with AAEEs, DAEEs, or allele CoEEs for specific genes in the brain.

The mechanisms regulating random monoallelic effects in cell lines are resistant to perturbations of DNA methylation and some histone modifications (Eckersley-Maslin et al., 2014; Gendrel et al., 2014). We found similar results for some of the DAEEs uncovered in our study (W.-C.H. and C.G., unpublished data). Nuclear architecture and chromatin looping may play important roles in regulating DAEEs versus allele CoEEs. CTCF and cohesin-mediated chromatin looping is required for genomic imprinting (Prickett et al., 2013, and references therein) and for the monoallelic expression of clustered protocadherins (Guo et al., 2012; Hirayama et al., 2012; Monahan et al., 2012). Additionally, allelic competition for enhancers contributes to the monoallelic expression of ORs (Markenscoff-Papadimitriou et al., 2014) and is also a plausible mechanism for DAEEs in our data. Overall, while genome architecture is cell type specific and probabilistic in nature (Cavalli and Misteli, 2013), little is known about allele architecture in specific brain cell populations. Advances in our ability to model sources of variance in RNA-seq data may further enhance the detection of non-genetic allelic effects using our approach, and new studies are needed to examine the temporal stability and mechanisms regulating the different allelic effects we uncovered.

The Functional Significance of Diverse Non-genetic Allelic Effects in the Brain

The functional significance of AAEEs, DAEEs, and CoEEs can be contemplated in the context of brain disorders and disease risk or as a biological mechanism of in vivo gene regulation. With respect to the former, mental illnesses are phenotypically variable and involve heterozygous variants and multiple genes; often, the same gene is implicated in different disorders (Gratten et al., 2014; McCarroll et al., 2014). Complex interactions between cellular non-genetic ASE effects and heterozygous genetic variants might be an important factor in shaping mental illness. Future studies are needed to determine whether the prevalence, identity, and location of monoallelic-MT cells in the brain can shape behavioral phenotypes in heterozygous mutant animals. Further, if environmental factors can change ASE effects, then complex interactions between environmental effects, allele genotype, and ASE effects may also have phenotypic consequences.

The potential in vivo functions of AAEEs, DAEEs, and CoEEs are unclear. Monoallelic expression is thought to come at an evolutionary cost to offspring fitness, since diploidy masks the effects of recessive mutations (Otto and Goldstein, 1992). We found that DRN autosomal genes with developmentally stable DAEEs are significantly enriched for DNA repair functions (data not shown). Genes with AAEEs or DAEEs are not expressed at lower levels than other autosomal genes and are not likely to play important roles in regulating gene dosage. In contrast, genes with allele CoEEs have elevated expression, indicating that promoting coordinated allelic expression could increase gene levels. The increased prevalence of DAEEs in the neonatal mouse DRN and cerebellum corresponds well with the peak of cell death in the developing brain (Ahern et al., 2013) and might increase genetic diversity for cell selection during periods of cell pruning. Other various functional roles for autosomal monoallelic effects have also been proposed (Chess, 2012), and in some cases, the phenomenon may simply reflect underlying genome architecture. Finally, our study uncovered significantly increased allele co-expression for X-linked genes in the juvenile and adult female brain compared to the neonatal brain or adult liver and muscle. Previous studies indicated a special role for X-linked genes in regulating brain function and cognition (Nguyen and Disteche, 2006, and references therein), and future studies that uncover the cellular basis of the developmental

of biallelic versus monoallelic brain cells is highly reproducible and consistent between individuals for each gene. *DEAF1* exhibits monoallelic expression in ~80% of positive cells, while *CNTNAP2* exhibits biallelic expression in ~70%–85% of positive cells. Dashed line indicates the percentage of cells with potential monoallelic expression due to cryosectioning artifacts.

(E) Comparison of the number of positive cells per cm^2 and the percentage of positive cells that are monoallelic between four different human brain regions for *CNTNAP2* and *DEAF1* (N = 5–12 cases per region). *CNTNAP2*+ cells are most prevalent in the ACC, while *DEAF1*+ cells are most prevalent in the OFC. The percentage of *CNTNAP2*+ cells that are monoallelic in the OFC is significantly greater than in the ACC or BA (one-way ANOVA with Tukey's HSD post-test). Significantly more monoallelic *CNTNAP2*+ cells were also observed in the DRN than in the ACC. Brain region differences in the percentage of monoallelic *DEAF1*+ cells were not observed. $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.0001$.

(F) Double-allele-specific in situ hybridization labeling for *CNTNAP2* or *DEAF1* (blue staining) and the neuron marker control gene *SYN2* (red staining) revealed subpopulations of monoallelic *CNTNAP2*+ neurons that exhibit biallelic *SYN2* expression in the human brain. In addition, subpopulations of biallelic *DEAF1*+ neurons were also identified in the human brain.

changes to X-linked allele co-expression in the female brain may reveal important new aspects of gene regulation on the X chromosome.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mouse Lines and Breeding
  - Macaque Brain Tissue
  - Human Brain Tissue
- METHOD DETAILS
  - Mouse and Primate Analysis of DAEEs
  - Analysis of Non-genetic Allelic Effects by RNA-Seq in the Mouse
  - Analysis of Non-genetic Allelic Effects by RNA-Seq in the Macaque
  - Estimation of the Effects of Genetic Variation on AAEEs, DAEEs, and Allele CoEEs in the Primate Brain
  - Examination of Genomic Imprinting in the Mouse and Primate Brain
  - Examination of the Conservation of Non-genetic Allelic Effects between Mice and Primates at the Gene Level
  - Nascent RNA and mRNA In Situ Hybridization Analysis
  - Mouse Tissue Processing
  - Human Tissue Processing
  - Pyrosequencing
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures, six tables, and supplemental text and can be found with this article online at http://dx.doi.org/10.1016/j.neuron.2017.01.033.

A video abstract is available at http://dx.doi.org/10.1016/j.neuron.2017.01.033#mmc9.

AUTHOR CONTRIBUTIONS

Conceptualization, W.-C.H., E.F., and C.G.; Methodology, W.-C.H., E.F., T.C., K.M.B., and C.G.; Investigation, W.-C.H., E.F., T.C., C.S.H., J.D.W., K.G., J.L.C., K.M.B., and C.G.; Software, E.F., W.-C.H., K.M.B., and C.G.; Formal Analysis, W.-C.H., E.F., T.C., and C.G.; Transgenic Mice, W.-C.H. and J.L.C.; Primate Tissue and Analysis, E.F., W.-C.H., J.D.W., C.S.H., and C.G.; Human Tissue Analysis, T.C., K.G., C.T., and C.G.; Data Visualization, W.-C.H., E.F., C.S.H., and C.G.; Writing, C.G., W.-C.H., and E.F.; Supervision, J.D.W., C.T., and C.G.; Funding, J.D.W. and C.G.

ACKNOWLEDGMENTS

We thank Drs. Suzi Mansour, Jared Rutter, and Brad Cairns and members of the Gregg lab for reviewing earlier versions of the manuscript, and we thank Drs. Li Zhang and Matthew Jorgensen for primate tissue collection. This work has been supported by NIH 1R01MH109577-01 (C.G.), a Simons Foundation Autism Research Initiative Explorer Award (C.G.), a University of Utah Seed Grant (C.G.), and NIH P40RR021380 (J.D.W.). C.G. is a New York Stem Cell Foundation Robertson Investigator. This research was supported by the New York Stem Cell Foundation.

Received: May 29, 2016
Revised: November 27, 2016
Accepted: January 30, 2017
Published: February 23, 2017

REFERENCES

Abelin, A.C.T., Marinov, G.K., Williams, B.A., McCue, K., and Wold, B.J. (2014). A ratiometric-based measure of gene co-expression. BMC Bioinformatics *15*, 331.

Ahern, T.H., Krug, S., Carr, A.V., Murray, E.K., Fitzpatrick, E., Bengston, L., McCutcheon, J., De Vries, G.J., and Forger, N.G. (2013). Cell death atlas of the postnatal mouse ventral forebrain and hypothalamus: effects of age and sex. J. Comp. Neurol. *521*, 2551–2569.

Alarcón, M., Abrahams, B.S., Stone, J.L., Duvall, J.A., Perederiy, J.V., Bomar, J.M., Sebat, J., Wigler, M., Martin, C.L., Ledbetter, D.H., et al. (2008). Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. Am. J. Hum. Genet. *82*, 150–159.

Amaral, D.G., Schumann, C.M., and Nordahl, C.W. (2008). Neuroanatomy of autism. Trends Neurosci. *31*, 137–145.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Arking, D.E., Cutler, D.J., Brune, C.W., Teslovich, T.M., West, K., Ikeda, M., Rea, A., Guy, M., Lin, S., Cook, E.H., and Chakravarti, A. (2008). A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism. Am. J. Hum. Genet. *82*, 160–164.

Babak, T., DeVeale, B., Tsang, E.K., Zhou, Y., Li, X., Smith, K.S., Kukurba, K.R., Zhang, R., Li, J.B., van der Kooy, D., et al. (2015). Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. Nat. Genet. *47*, 544–549.

Bauer-Mehren, A., Rautschka, M., Sanz, F., and Furlong, L.I. (2010). DisGeNET: a Cytoscape plugin to visualize, integrate, search and analyze gene-disease networks. Bioinformatics *26*, 2924–2926.

Bonthuis, P.J., Huang, W.-C., Stacher Hörndli, C.N., Ferris, E., Cheng, T., and Gregg, C. (2015). Noncanonical genomic imprinting effects in offspring. Cell Rep. *12*, 979–991.

Borel, C., Ferreira, P.G., Santoni, F., Delaneau, O., Fort, A., Popadin, K.Y., Garieri, M., Falconnet, E., Ribaux, P., Guipponi, M., et al. (2015). Biased allelic expression in human primary fibroblast single cells. Am. J. Hum. Genet. *96*, 70–80.

Bullard, J.H., Purdom, E., Hansen, K.D., and Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics *11*, 94.

Cavalli, G., and Misteli, T. (2013). Functional implications of genome topology. Nat. Struct. Mol. Biol. *20*, 290–299.

Chess, A. (2012). Mechanisms and consequences of widespread random monoallelic expression. Nat. Rev. Genet. *13*, 421–428.

Chess, A. (2016). Monoallelic gene expression in mammals. Annu. Rev. Genet. *50*, 317–327.

Crowley, J.J., Zhabotynsky, V., Sun, W., Huang, S., Pakatci, I.K., Kim, Y., Wang, J.R., Morgan, A.P., Calaway, J.D., Aylor, D.L., et al. (2015). Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. Nat. Genet. *47*, 353–360.

Deng, X., Berletch, J.B., Nguyen, D.K., and Disteche, C.M. (2014a). X chromosome regulation: diverse patterns in development, tissues and disease. Nat. Rev. Genet. *15*, 367–378.

Deng, Q., Ramsköld, D., Reinius, B., and Sandberg, R. (2014b). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science *343*, 193–196.

Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. Nature *518*, 331–336.

Eckersley-Maslin, M.A., Thybert, D., Bergmann, J.H., Marioni, J.C., Flicek, P., and Spector, D.L. (2014). Random monoallelic gene expression increases upon embryonic stem cell differentiation. Dev. Cell *28*, 351–365.

Gendrel, A.-V., Attia, M., Chen, C.-J., Diabangouaya, P., Servant, N., Barillot, E., and Heard, E. (2014). Developmental dynamics and disease potential of random monoallelic gene expression. Dev. Cell *28*, 366–380.

Gimelbrant, A., Hutchinson, J.N., Thompson, B.R., and Chess, A. (2007). Widespread monoallelic expression on human autosomes. Science *318*, 1136–1140.

Godin, R.E., Takaesu, N.T., Robertson, E.J., and Dudley, A.T. (1998). Regulation of BMP7 expression during kidney development. Development *125*, 3473–3482.

Gratten, J., Wray, N.R., Keller, M.C., and Visscher, P.M. (2014). Large-scale genomics unveils the genetic architecture of psychiatric disorders. Nat. Neurosci. *17*, 782–790.

Guo, Y., Monahan, K., Wu, H., Gertz, J., Varley, K.E., Li, W., Myers, R.M., Maniatis, T., and Wu, Q. (2012). CTCF/cohesin-mediated DNA looping is required for protocadherin α promoter choice. Proc. Natl. Acad. Sci. USA *109*, 21081–21086.

He, B., Chen, C., Teng, L., and Tan, K. (2014). Global view of enhancer-promoter interactome in human cells. Proc. Natl. Acad. Sci. USA *111*, E2191–E2199.

Heinz, S., Romanoski, C.E., Benner, C., Allison, K.A., Kaikkonen, M.U., Orozco, L.D., and Glass, C.K. (2013). Effect of natural genetic variation on enhancer selection and function. Nature *503*, 487–492.

Higashino, A., Sakate, R., Kameoka, Y., Takahashi, I., Hirata, M., Tanuma, R., Masui, T., Yasutomi, Y., and Osada, N. (2012). Whole-genome sequencing and analysis of the Malaysian cynomolgus macaque (Macaca fascicularis) genome. Genome Biol. *13*, R58.

Hirano, K., Kaneko, R., Izawa, T., Kawaguchi, M., Kitsukawa, T., and Yagi, T. (2012). Single-neuron diversity generated by Protocadherin-β cluster in mouse central and peripheral nervous systems. Front. Mol. Neurosci. *5*, 90.

Hirayama, T., Tarusawa, E., Yoshimura, Y., Galjart, N., and Yagi, T. (2012). CTCF is required for neural development and stochastic expression of clustered Pcdh genes in neurons. Cell Rep. *2*, 345–357.

Jeffries, A.R., Perfect, L.W., Ledderose, J., Schalkwyk, L.C., Bray, N.J., Mill, J., and Price, J. (2012). Stochastic choice of allelic expression in human neural stem cells. Stem Cells *30*, 1938–1947.

Jeffries, A.R., Collier, D.A., Vassos, E., Curran, S., Ogilvie, C.M., and Price, J. (2013). Random or stochastic monoallelic expressed genes are enriched for neurodevelopmental disorder candidate genes. PLoS ONE *8*, e85093.

Kasowski, M., Kyriazopoulou-Panagiotopoulou, S., Grubert, F., Zaugg, J.B., Kundaje, A., Liu, Y., Boyle, A.P., Zhang, Q.C., Zakharia, F., Spacek, D.V., et al. (2013). Extensive variation in chromatin states across humans. Science *342*, 750–752.

Kilpinen, H., Waszak, S.M., Gschwind, A.R., Raghav, S.K., Witwicki, R.M., Orioli, A., Migliavacca, E., Wiederkehr, M., Gutierrez-Arcelus, M., Panousis, N.I., et al. (2013). Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. Science *342*, 744–747.

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.

Lawson, K.A., Dunn, N.R., Roelen, B.A., Zeinstra, L.M., Davis, A.M., Wright, C.V., Korving, J.P., and Hogan, B.L. (1999). Bmp4 is required for the generation of primordial germ cells in the mouse embryo. Genes Dev. *13*, 424–436.

Leung, D., Jung, I., Rajagopal, N., Schmitt, A., Selvaraj, S., Lee, A.Y., Yen, C.-A., Lin, S., Lin, Y., Qiu, Y., et al. (2015). Integrative analysis of haplotype-resolved epigenomes across human tissues. Nature *518*, 350–354.

Marinov, G.K., Williams, B.A., McCue, K., Schroth, G.P., Gertz, J., Myers, R.M., and Wold, B.J. (2014). From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. Genome Res. *24*, 496–510.

Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. *18*, 1509–1517.

Markenscoff-Papadimitriou, E., Allen, W.E., Colquitt, B.M., Goh, T., Murphy, K.K., Monahan, K., Mosley, C.P., Ahituv, N., and Lomvardas, S. (2014). Enhancer interaction networks as a means for singular olfactory receptor expression. Cell *159*, 543–557.

McCarroll, S.A., Feng, G., and Hyman, S.E. (2014). Genome-scale neurogenetics: methodology and meaning. Nat. Neurosci. *17*, 756–763.

Mekhoubad, S., Bock, C., de Boer, A.S., Kiskinis, E., Meissner, A., and Eggan, K. (2012). Erosion of dosage compensation impacts human iPSC disease modeling. Cell Stem Cell *10*, 595–609.

Monahan, K., Rudnick, N.D., Kehayova, P.D., Pauli, F., Newberry, K.M., Myers, R.M., and Maniatis, T. (2012). Role of CCCTC binding factor (CTCF) and cohesin in the generation of single-cell diversity of protocadherin-α gene expression. Proc. Natl. Acad. Sci. USA *109*, 9125–9130.

Nag, A., Savova, V., Fung, H.-L., Miron, A., Yuan, G.-C., Zhang, K., and Gimelbrant, A.A. (2013). Chromatin signature of widespread monoallelic expression. eLife *2*, e01256.

Nazor, K.L., Altun, G., Lynch, C., Tran, H., Harness, J.V., Slavin, I., Garitaonandia, I., Müller, F.-J., Wang, Y.-C., Boscolo, F.S., et al. (2012). Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. Cell Stem Cell *10*, 620–634.

Nguyen, D.K., and Disteche, C.M. (2006). High expression of the mammalian X chromosome in brain. Brain Res. *1126*, 46–49.

Oberg, A.L., Bot, B.M., Grill, D.E., Poland, G.A., and Therneau, T.M. (2012). Technical and biological variance structure in mRNA-Seq data: life in the real world. BMC Genomics *13*, 304.

Otto, S.P., and Goldstein, D.B. (1992). Recombination and the evolution of diploidy. Genetics *131*, 745–751.

Perez, J.D., Rubinstein, N.D., Fernandez, D.E., Santoro, S.W., Needleman, L.A., Ho-Shing, O., Choi, J.J., Zirlinger, M., Chen, S.-K., Liu, J.S., and Dulac, C. (2015). Quantitative and functional interrogation of parent-of-origin allelic expression biases in the brain. eLife *4*, e07860.

Peters, J. (2014). The role of genomic imprinting in biology and disease: an expanding view. Nat. Rev. Genet. *15*, 517–530.

Pinter, S.F., Colognori, D., Beliveau, B.J., Sadreyev, R.I., Payer, B., Yildirim, E., Wu, C.-T., and Lee, J.T. (2015). Allelic imbalance is a prevalent and tissue-specific feature of the mouse transcriptome. Genetics *200*, 537–549.

Piras, V., and Selvarajoo, K. (2015). The reduction of gene expression variability from single cells to populations follows simple statistical laws. Genomics *105*, 137–144.

Prickett, A.R., Barkas, N., McCole, R.B., Hughes, S., Amante, S.M., Schulz, R., and Oakey, R.J. (2013). Genome-wide and parental allele-specific analysis of CTCF and cohesin DNA binding in mouse brain reveals a tissue-specific binding pattern and an association with imprinted differentially methylated regions. Genome Res. *23*, 1624–1635.

Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. Cell *135*, 216–226.

Rajab, A., Schuelke, M., Gill, E., Zwirner, A., Seifert, F., Morales Gonzalez, S., and Knierim, E. (2015). Recessive DEAF1 mutation associates with autism, intellectual disability, basal ganglia dysfunction and epilepsy. J. Med. Genet. *52*, 607–611.

Reinius, B., Mold, J.E., Ramsköld, D., Deng, Q., Johnsson, P., Michaëlsson, J., Frisén, J., and Sandberg, R. (2016). Analysis of allelic expression patterns in clonal somatic cells by single-cell RNA-seq. Nat. Genet. *48*, 1430–1435.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Savova, V., Chun, S., Sohail, M., McCole, R.B., Witwicki, R., Gai, L., Lenz, T.L., Wu, C.-T., Sunyaev, S.R., and Gimelbrant, A.A. (2016). Genes with monoallelic expression contribute disproportionately to genetic diversity in humans. Nat. Genet. *48*, 231–237.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.

Segklia, A., Seuntjens, E., Elkouris, M., Tsalavos, S., Stappers, E., Mitsiadis, T.A., Huylebroeck, D., Remboutsika, E., and Graf, D. (2012). Bmp7 regulates the survival, proliferation, and neurogenic properties of neural progenitor cells during corticogenesis in the mouse. PLoS ONE *7*, e34088.

Soiza-Reilly, M., and Commons, K.G. (2014). Unraveling the architecture of the dorsal raphe synaptic neuropil using high-resolution neuroanatomy. Front Neural Circuits *8*, 105.

Stadtfeld, M., Apostolou, E., Ferrari, F., Choi, J., Walsh, R.M., Chen, T., Ooi, S.S.K., Kim, S.Y., Bestor, T.H., Shioda, T., et al. (2012). Ascorbic acid prevents loss of Dlk1-Dio3 imprinting and facilitates generation of all-iPS cell mice from terminally differentiated B cells. Nat. Genet. *44*, 398–405, S1–S2.

Vulto-van Silfhout, A.T., Rajamanickam, S., Jensik, P.J., Vergult, S., de Rocker, N., Newhall, K.J., Raghavan, R., Reardon, S.N., Jarrett, K., McIntyre, T., et al. (2014). Mutations affecting the SAND domain of DEAF1 cause intellectual disability with severe speech impairment and behavioral problems. Am. J. Hum. Genet. *94*, 649–661.

Yan, G., Zhang, G., Fang, X., Zhang, Y., Li, C., Ling, F., Cooper, D.N., Li, Q., Li, Y., van Gool, A.J., et al. (2011). Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaques. Nat. Biotechnol. *29*, 1019–1023.

Yang, F., Babak, T., Shendure, J., and Disteche, C.M. (2010). Global survey of escape from X inactivation by RNA-sequencing in mouse. Genome Res. *20*, 614–622.

Zwemer, L.M., Zak, A., Thompson, B.R., Kirby, A., Daly, M.J., Chess, A., and Gimelbrant, A.A. (2012). Autosomal monoallelic expression in the mouse. Genome Biol. *13*, R10.

CellPress

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Primate brain and blood samples	Wake Forest School of Medicine	NA
Human brain tissue	Dallas Brain Collection and NIH Neurobiobank	NA
Critical Commercial Assays		
RNAscope 2.5 HD Detection Reagent - RED	Advanced Cell Diagnostics	Cat#322360
RNAscope 2-plex Detection Reagents	Advanced Cell Diagnostics	Cat#320701
PyroMark Gold Q24 Reagents	QIAGEN	Cat#970802
RNeasy Micro Kit	QIAGEN	Cat#74004
Deposited Data		
Raw and Analyzed data	This paper	GEO: GSE93788
Raw data of RNA-seq on mouse ARN, DRN, liver and muscle	Bonthuis et al. (2015)	GEO: GSE70484
Allele count table for P8 and P60 cerebellum comparison	Perez et al. (2015)	GEO: GSE67556
Published random monoallelic autosomal gene lists	Gendrel et al. (2014), Eckersley-Maslin et al. (2014), Zwemer et al. (2012), Reinius et al. (2016), Savova et al. (2016)	Sources listed in Table S3
Mouse reference genome (NCBI37/mm9)	Genome Reference Consortium	https://www.ncbi.nlm.nih.gov/grc/mouse
Experimental Models: Organisms/Strains		
Mouse: $Bmp7^{LacZ/+}$ / $Bmp4^{LacZ/+}$	Godin et al., 1998; Lawson et al., 1999	N/A
Mouse: B6N(Cg)-Adora2b < tm1.1(KOMP)V1cg > /J	The Jackson Laboratory	#018592
Mouse: C57BL/6J	The Jackson Laboratory	#000664
Mouse: CAST/EiJ	The Jackson Laboratory	#000928
Sequence-Based Reagents		
Primers for pyrosequencing	This paper	Table S4
Ecoli-LacZ-C2	Advanced Cell Diagnostics	Cat#313451-C2
Hs-CNTNAP2-intron	Advanced Cell Diagnostics	Cat#435841
Hs-DEAF1-intron	Advanced Cell Diagnostics	Cat#435851
Mm-Adora2b	Advanced Cell Diagnostics	Cat#445281
Mm-Adora2b-Intron	Advanced Cell Diagnostics	Cat#435861
Mm-Ank2-Intron4	Advanced Cell Diagnostics	Cat#423941
Mm-Atp8a1-Intron16	Advanced Cell Diagnostics	Cat#423951
Mm-Bmp4	Advanced Cell Diagnostics	Cat#401301
Mm-Bmp4-Intron3	Advanced Cell Diagnostics	Cat#449861
Mm-Bmp7	Advanced Cell Diagnostics	Cat#407901
Mm-Bmp7-Intron5	Advanced Cell Diagnostics	Cat#444521
Mm-Cacng2-Intron1	Advanced Cell Diagnostics	Cat#478221
Mm-Mtap1b-Intron2	Advanced Cell Diagnostics	Cat#439861
Mm-Nr1d1-Intron1	Advanced Cell Diagnostics	Cat#478231
Mm-Oxtr-Intron2	Advanced Cell Diagnostics	Cat#478241
Mm-Pld3-Intron9	Advanced Cell Diagnostics	Cat#423961
Mm-Syn2-Intron-C2	Advanced Cell Diagnostics	Cat#436631-C2

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
R	R Core Team (2016)	https://www.r-project.org/
ImageJ	Schneider et al. (2012)	https://imagej.nih.gov/ij/
PyroMark Q24 Software	QIAGEN	v2.0.6.20
PyroMark Assay Design SW	QIAGEN	v2.0
Prism	GraphPad	http://www.graphpad.com
Novoalign	Novocraft	http://www.novocraft.com/
EdgeR	Robinson et al. (2010)	https://bioconductor.org/packages/release/bioc/html/edgeR.html

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dr. Christopher Gregg (chris.gregg@neuro.utah.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mouse Lines and Breeding

All mouse experiments were performed in accordance with protocols approved by the University of Utah Institutional Animal Care and Use Committee. B6, Cast and *Adora2b*$^{+/LacZ}$ mice were obtained from the Jackson Laboratory. *Bmp7*$^{+/LacZ}$ and *Bmp4*$^{+/LacZ}$ mice have been previously described (Godin et al., 1998; Lawson et al., 1999). P0 is the day of birth. Adult mice in our study are 8-12 weeks of age. Female mice are analyzed in all cases, so that random X inactivation serves as an internal control. Mice were maintained on 12 hr light/dark cycle and given water and food ad libitum (Harlan Tecklad rodent diet 2920X; Madison, WI.). Cage bedding was Paperchip bedding (Shepherd Specialty Paper) and animals were housed with sex-matched littermates (2-5 mice per cage) and weaned at P21.

Macaque Brain Tissue

All primate experiments were performed in accordance with protocols approved by the Wake Forest University School of Medicine Institutional Animal Care and Use Committee. DRN brain tissue was obtained from juvenile cynomolgus macaque females for RNA-seq analysis by micro-dissection from frozen brain tissue slabs. DRN micro-dissections included the ventral PAG region. Blood samples were obtained from the parents for each daughter for DNA extraction and genome sequencing. Detailed information for each sample is provided in Table S4.

Human Brain Tissue

Human adult female fresh frozen brain tissue samples for target regions were obtained from control cases in the Dallas Brain Collection and the NIH Neurobiobank. Reported RIN values (RNA Integrity Number) for analyzed brains was > 7. Neuropathic analysis was normal for each case. Age ranged from 53 to 83 years. Tissue obtained from the NIH Neurobiobank came from the Human Brain and Spinal Fluid Resource Center, VA West Los Angeles Heathcare Center, Los Angeles, CA 90073, which is sponsored by NINDS/NIMH, National Multiple Sclerosis Society and the Department of Veterans Affairs. Post-mortem human brain tissue obtained from the UT Neuropsychiatry Research Program (Dallas Brain Collection-DBC) at UT Southwestern Medical Center was collected after consent from the deceased/donor's legal next of kin (NOK) was received, along with permission to access medical records and to conduct a direct telephone interview with the NOK or a primary caregiver. All clinical and medical information obtained for each donor is reviewed by a research panel, composed of at least three research and/or clinical psychiatrists and psychologists, after which a consensus diagnoses for each case is made using DSM-IV criteria. Blood and vitreous toxicology screens for drugs of abuse, alcohol, and prescription drugs, are conducted for each donor subject when possible. Collection of post-mortem human brain tissue is approved by the University of Texas Southwestern Medical Center Institutional Review Board [IRB#102010-053].

METHOD DETAILS

Mouse and Primate Analysis of DAEEs

RNA-seq in reciprocal F1 Cast × B6 hybrid mice was performed as detailed previously (Bonthuis et al., 2015). For the macaque DRN, RNA was extracted using the QIAGEN RNeasy Kit and whole transcriptome sequencing was performed using the RiboZero TruSeq Stranded Total RNA Library Prep Kit (Illumina). Mouse and primate DRN micro-dissections included the ventral PAG region. Whole

genome sequencing of the parents was performed using the TruSeq DNA PCR-free Library Prep Kit (Illumina). Primate transcriptome and genome sequencing involved performing 100bp paired-end read sequencing on a HiSeq 2500 (Illumina). Data analysis was performed using GATK and custom software as detailed below. Metrics for all genomics experiments are presented in Table S4. The number of biological replicates for each mouse tissue RNA-seq experiment are: P5 female DRN (n = 14 (7 replicates per F1cb/F1bc cross)), P15 female DRN (n = 14 (7 replicates per F1cb/F1bc cross)), adult female DRN (n = 18 (9 replicates per F1cb/F1bc cross)), arcuate nucleus (ARN) of the hypothalamus (n = 18 (9 replicates per F1cb/F1bc cross)), adult female liver (endoderm-derived, n = 16 (8 per F1cb/F1bc cross)) and adult female skeletal muscle (mesoderm-derived, n = 16 (8 per F1cb/F1bc cross)), P8 cerebellum (n = 12 (6 per F1cb/F1bc cross)), and P60 cerebellum (n = 12 (6 per F1cb/F1bc cross)).

Analysis of Non-genetic Allelic Effects by RNA-Seq in the Mouse

The DRN (dorsal raphe nucleus) was micro-dissected from P5, P15 and adult female F1i and F1r hybrid mice, as described previously (Bonthuis et al., 2015). The DRN dissections include portions of the ventral periaqueductal gray. RNA was extracted using the RNAeasy Micro Kit (QIAGEN). RNA was pooled from 4-10 offspring from at least three different litters to provide ~3 ug of total RNA for each biological replicate. Samples were prepared by the University of Utah High-Throughput Genomics core facility with Illumina TruSeq RNA Sample Prep v2 with oligo dT selection for HiSeq 59 Cycle Single-Read Sequencing v4 and sequenced on an Illumina HiSeq 2500 (Illumina). For adults, 9 F1i and 9 F1r biological replicates were performed, for P15 and P5, 7 F1i and 7 F1r biological replicates were performed. Datasets for the adult female ARN, liver and muscle have been previously described (Bonthuis et al., 2015).

To examine allele co-expression in mice, Cast and B6 allele read counts were normalized to counts per million (CPM) using EdgeR (Robinson et al., 2010). Normalization factors were calculated for the sum of both alleles and then applied to the alleles counts separately. The mouse cerebellum allelic RNA-seq dataset was obtained from Perez et al. (2015)'s published dataset (GEO: GSE67556). To remove cross effects in our allele expression data, the allele expression data was centered for each gene by subtracting the mean Cast and B6 allele expression level from the actual Cast and B6 allele expression data, respectively, for each cross. Centered values were whitened by dividing by the standard deviation of parental allele expression. After centering and whitening (Figure 1B), Pearson correlation coefficients (r_a values) were computed for maternal and paternal allele co-expression for each gene.

Genes differ in terms of expression level and biological variance characteristics, and the allele expression correlation values (r_a values) for genes with low expression and little variance will be more susceptible to the effects of technical noise than genes with high expression and high biological variance. To address this problem, we devised an empirical Bayesian approach to estimate the 95% confidence interval (CIs) for the ground-truth, biological r_a value for each gene (r_{ab}). For each dataset, we categorized genes according to their expression level characteristics by defining 100 different expression level quantiles in our RNA-seq data (Figure S1A). Next, we categorized genes according to their variance characteristics by calculating the biological coefficient of variance (BCV) from the RNA-seq replicates for each gene based on the SNP aligning reads using the edgeR statistical package (Robinson et al., 2010). From the BCV for each gene in the data, we defined 20 quantiles of genes based on the BCV values for each dataset (Figure S1B). Therefore, for each age or tissue type RNA-seq dataset, we defined a total of 2000 different expression level × BCV conditions (100 × 20) that represent the different expression characteristics of the genes.

For each of these 2000 different conditions, we modeled the impact of biological variance (gamma distribution) and RNA-seq technical noise (Poisson distribution) on ground-truth r_{ab} values ranging from −1 to +1 (in 0.05 increments). In our approach, we simulated maternal and paternal reads for each of the RNA-seq replicates for each r_{ab} value based on the gamma distribution with mean expression and BCV characteristics that are appropriate for the expression level × BCV condition to be modeled. We then simulated the effects of RNA-seq technical noise by performing Poisson resampling of the initially simulated reads. These two steps are iterated 10,000 times, which generates a distribution of observed r_a values for each one of the ground-truth r_{ab} values modeled (from −1 to +1 in 0.05 increments). We perform this entire modeling procedure for all 2000 expression × BCV conditions in the data. Our approach is based on studies that showed that technical variance in RNA-seq data is accurately modeled by the Poisson distribution (Bullard et al., 2010; Marioni et al., 2008; Oberg et al., 2012), and that combined biological + technical variance can be modeled using the negative binomial distribution (Anders and Huber, 2010; Oberg et al., 2012; Robinson et al., 2010). Similar to our approach, the negative binomial is a gamma-Poisson mixture distribution.

The goal of our modeling is to be able to define the 95% CI for the ground-truth r_{ab} statistic for any observed r_a value in our RNA-seq data. In order to achieve this goal, we tallied the number of times each r_a value was observed from the modeling results for each ground-truth r_{ab}. These results were assembled into frequency tables in which the x axis is the ground-truth r_{ab} value and the y axis is the observed r_a values derived from the modeling. One table was assembled for each of the 2000 expression × BCV conditions. Thus, for any observed r_a value in our RNA-seq data, we look up the table with the corresponding expression × BCV characteristics of the gene, identify the row for the observed r_a value and extract the data for the entire row, thereby providing the frequency distribution of ground-truth r_{ab} values from which we define the r_{ab} 95% CI for the gene.

In the final step of our procedure, we weight the frequency distributions in the tables based on an empirically determined prior probability distribution. We expect that most X-linked genes will exhibit low r_{ab} values due to random X inactivation in the female brain, and most autosomal genes will exhibit relatively high r_{ab} values due to biallelic expression. Indeed, the observed r_a values in our RNA-seq results are consistent with this expectation, since the distribution for autosomal genes peaks near +1.0, while the distribution for

X-linked genes peaks near 0, or for some datasets peak at a negative r_a value (Figure S1C). To reflect this prior knowledge, we re-weight the modeled data tables according to a Beta distribution, ranging from −1 to +1 (in 0.05 bins), with parameters derived from the distribution of observed r_a values for autosomal or X-linked genes (Figure S1D). This prior has little impact on most of the data, but changes the CIs for autosomal and X-linked genes with the lowest expression and variance characteristics by shifting the CIs toward the overall trend in the data (Figure S1D). We found that we cannot reliably model genes with a mean SNP-aligning read expression level of 0.2 counts per million (CPM) or less, and removed these genes. In addition, in the mouse studies, genes that monoallelically express Cast or B6 alleles due to genetic effects were filtered from the datasets.

Analysis of Non-genetic Allelic Effects by RNA-Seq in the Macaque

For the macaque study, RNA was extracted from micro-dissected fresh frozen brain tissue sections of the macaque DRN and ventral periaqueductal gray region using the QIAGEN RNeasy Kit. Whole transcriptome sequencing was performed using the RiboZero TruSeq Stranded Total RNA Library Prep Kit (Illumina). For parental genome analysis, whole genome sequencing of the parents was performed on DNA extracted from blood samples using the TruSeq DNA PCR-free Library Prep Kit (Illumina). Metrics for all genomics experiments are presented in Table S5. Primate transcriptome and genome sequencing involved performing 100bp paired-end read sequencing on a HiSeq 2500 (Illumina). Data analysis was performed using GATK and custom software as detailed below.

The genomes of the parents in the Crab-eating macaque (*Macaca fascicularis*) study were sequenced and aligned to the macFas5 genome using Novoalign (Novocraft Technologies). These parental genomes were genotyped using the GATK Unified Genotyper. The resulting variant calls were used to identify SNPs that are potentially heterozygous in the offspring. This information was incorporated into individualized transcriptome indices in which IUPAC codes indicated the ambiguity of SNPs that might be heterozygous in the offspring allowing for an unbiased daughter transcriptome alignment, which was also performed with Novoalign. Daughter RNA-seq reads containing heterozygous SNPs were counted as maternal or paternal reads using a custom pysam-based python script that accounts for the parental genotypes and the daughter genotype apparent in the transcriptome alignment. Maternal and Paternal distinguishing reads for 838 genes with distinguishing SNPs in all 10 daughters were summed for each gene and CPM normalized using EdgeR (Robinson et al., 2010). The allele expression levels were centered by subtracting the mean expression level. Pearson's correlation values were calculated across the 10 daughter macaques and confidence intervals for each gene were estimated based on gamma-Poisson resampling of the data as detailed above for the mouse data.

Estimation of the Effects of Genetic Variation on AAEEs, DAEEs, and Allele CoEEs in the Primate Brain

To estimate the impact of genetic variation in *cis*-regulatory regions for the primate data, we performed the following analysis. To begin, we sought to estimate the proportion of genes with genetic effects on allelic expression in each daughter. Our approach takes advantage of data in F1cb and F1bc F1 hybrid mice, for which we know the identity of the genes with genetic allelic effects. In these mice, the genes are impacted in the same manner in each individual, since the animals are derived from crosses of inbred lines and there is very little genetic variation. Previously, we identified all genes in F1cb and F1bc mice with statistically significant Cast versus B6 allele expression biases in the female mouse DRN (Bonthuis et al., 2015). We determined that 68% of all expressed genes (10,421 genes in total) in the F1cb and F1bc mouse DRN have a significant strain genetic effect (Cast or B6 allele bias). We sought to express this effect as a function of the amount of genetic variation in putative functional non-coding regulatory regions in the genome, in order to use this ratio to estimate the proportion of genes impacted by genetic effects in each primate daughter. Promoter regions are non-coding regulatory elements that are often defined as the region 2kbp upstream of a transcriptional start site (He et al., 2014) and can be defined in both mice and macaques. In contrast, enhancers are not well annotated in cynomolgus macaques; therefore, for the purposes of estimating genetic variation in noncoding regulatory elements in each species, we focused on promoter regions in the mouse and macaque genome, defined as the sequence 2 kpb upstream of a transcriptional start site. We then analyzed genetic variation within these regions in F1cb/F1bc mice and in each macaque daughter.

For the mice, Cast versus B6 genome sequence variants were obtained from the Sanger Mouse Genome Project *vcf* files. From these files, we identified all Cast/B6 variant sites and intersected the sites with the bedfile of mouse promoter regions. The analysis revealed a variant rate of 8.11 variants per kb in putative regulatory elements (VR_{reg}) for the Cast versus B6 genome contrast. We can therefore express the percentage of genes with genetic allelic effects (P_{GE}) in the mouse as a function of the VR_{reg},

$$68\%\ P_{GE(Mouse)}$$

$$8.11\ VR_{reg}$$

With this data for the mouse, we can use the maternal and paternal genome sequences for each primate daughter to determine VR_{reg} and then estimate the P_{GE} for each primate daughter. Our approach assumes that species/individuals with more genetic variation in promoter regions throughout the genome will have more genes with genetic allelic effects as compared to species/individuals with less genetic variation. Based on this assumption, and our above analysis of Cast × B6 F1 hybrid mice, we estimated the percentage of genes that are expected to have significant genetic allelic effects in each of the macaque daughters, as follows. We used the genome sequence data we generated for each of the macaque mothers and fathers to compute the VR_{reg} for each daughter

(Table S5). From this value and the mouse data, we estimate the percentage of genes with genetic allelic expression effects (P_{GE}) in each of the ten daughters as follows:

$$\frac{P_{GE(Mouse)}}{VR_{reg(Mouse)}} = \frac{P_{GE(Primate)}}{VR_{reg(Primate)}}.$$

Next, we estimate the probability that a single gene will have a genetic effect (P_{GE}) in each daughter (Table S5). Future studies that map various regulatory elements in the mouse and macaque genomes may be able to refine this estimation, which is currently based on promoter regions only.

Having estimated the proportion of genes with genetic allelic effects in each daughter, we next sought to estimate the magnitude of the genetic allele expression bias exhibited by each impacted primate gene. This estimation is also derived from the mouse data. As noted above, we have identified genes in F1cb and F1bc mice with significant Cast or B6 allele expression biases due to cis-regulatory genetic differences between Cast and B6 mice (Bonthuis et al., 2015) (FDR 5%, data presented in Table S6). In addition, we determined the fold Cast or B6 allele expression bias due to genetic effects for each significantly impacted gene (Table S6, see log2 fold allele bias column). We use these results as a representation of the distribution of the allelic effect sizes expected in the primate due to genetic variation. Thus, by working from mouse data for which the allelic effects are well controlled and easily interpreted, we are able to estimate both the probability of a genetic effect for a given gene in each primate daughter and the magnitude of the resulting allelic expression bias. As detailed below, we introduce these parameters into our empirical Bayesian modeling procedure to estimate the impact of genetic variation on our confidence on the underlying non-genetic r_{ab} value for each gene in our primate data.

Similar to our modeling approach to define the r_{ab} 95% CI for the mouse, our statistical modeling for the primate involves first seeding reads for a non-genetic, ground-truth allele correlation, r_{ab}, which is performed based on the gamma distribution to simulate biological variance. The reads are seeded with a mean and variance determined from the expression level and BCV computed from the primate gene expression data. Reads are seeded for ten replicates representing each of the ten primate daughters. Next, we introduce genetic effects into the simulated data. For each of the ten daughters, we randomly assign each daughter a score of 1 (genetic effect) or 0 (non-genetic effect) using a binomial distribution weighted according to the estimated probability of a gene having a genetic effect in that daughter (Table S5). If the daughter draws a genetic effect, one allele is randomly selected and multiplied by a fold allele expression bias drawn at random from the distribution of Cast versus B6 allele expression biases in the mouse, thereby introducing the estimated genetic effect into the data. This procedure is performed for each of the ten primate daughters. Once the genetic effects have been introduced, Poisson resampling is performed to model the effects of RNA-seq technical variance and the observed r_a value is calculated from the data. This entire procedure is iterated 10,000 times for each gene, generating a distribution of observed r_a values that reflects the effects of biological variation, genetic variation and technical noise in the data. This modeling procedure is repeated for all ground-truth, non-genetic effect r_{ab} values from −1 to +1 in 0.05 increments. The method provides a look up table of observed r_a values (y axis) and corresponding ground-truth, non-genetic r_{ab} values (x axis). For the primate, since we only have 838 genes to model, rather than working from expression level and estimated BCV quantiles as was done in the mouse, we generate one lookup table for every gene based on the mean expression level and BCV for that gene. With these tables, we define the 95% CI for the ground-truth, non-genetic r_{ab} for each gene in the primate.

Examination of Genomic Imprinting in the Mouse and Primate Brain

Genomic imprinting analysis was performed using the generalized linear modeling function in edgeR to test for a main effect of the parental origin of the allele (mice: (read count ~strain-of-origin + parent-of-origin); primate: (allele read count ~parent-of-origin)), as previously described (Bonthuis et al., 2015). P values were adjusted for multiple testing using the BH method. For mice, the false discovery rate (FDR) threshold for imprinting effects was set to 1%. For the primate datasets, we did not identify any genes with significant imprinting effects at either the 1% or the 5% FDR thresholds. Metrics for all genomics datasets are in Table S4.

Examination of the Conservation of Non-genetic Allelic Effects between Mice and Primates at the Gene Level

For each gene, we estimate the ground-truth, non-genetic r_{ab} 95% CI for allelic effects and from our modeling tables for CI estimation we are able to define the median r_{ab} value for each gene. The median r_{ab} value is the best single value to represent the underlying non-genetic allelic effect for a gene. Therefore, for primate genes analyzed for allelic effects we compare the juvenile primate median r_{ab} value to the juvenile mouse median r_{ab} value for the mouse ortholog of the gene. If the effects are well conserved between the two species, we expect the median r_{ab} values to be correlated after differences in expression level have been controlled for. To do this comparison, we calculated the partial correlation coefficient for the non-parametric Spearman Rank statistic using the pcor.test function in R.

Nascent RNA and mRNA In Situ Hybridization Analysis

Nascent RNA in situ hybridization analysis was performed using RNAscope probes and HD 2.5 staining kit (Advanced Cell Diagnostics) as previously described (Bonthuis et al., 2015). RNAscope in situ hybridization for mRNA was performed using the

RNAscope 2-plex kit according to the manufacturers instructions. The manufacturers positive control probe was used to evaluate RNA integrity in human brain tissue samples and samples with poor staining were not analyzed.

Mouse Tissue Processing

Wild-type or mutant adult female mice were perfused by 4% paraformaldehyde (PFA) and then fixed in 4% PFA overnight at 4°C. The brain was immersed in 10% sucrose at 4°C for 24 hr. This step was repeated with 20% and 30% sucrose. The brain was embedded in *Optimal Cutting Temperature* compound (OCT) and stored at −80°C. ARN and DRN 14 μm sections were collected using a cryostat. RNA signal was detected following the RNAscope Chromogenic Assay Sample Preparation for Fixed Frozen Tissue (Part 1) technical note and RNAscope 2-Plex or HD Kit (Part 2). E14.5 brains were dissected from the pregnant mouse and then fixed in 4% PFA overnight at 4°C. The steps of sucrose dehydration, cryosection, and staining are described above. P5 brain tissues were embedded directly after dissection and 14 μm sections were collected by cryosection. The staining was performed following the RNAscope Chromogenic Assay Sample Preparation for Fresh Frozen Tissue (Part 1) and following the instructions for the RNAscope 2-Plex Kit (Part 2). Tissue sections were imaged with a Zeiss ApoTome2 microscope using a 40x oil objective. Three images were taken per section from at least three sections per animal for nascent RNA detection. Four images were taken per section from two sections per animal for wild-type/LacZ allele detection in heterozygous mice. Cells were counted using Cell Counter function in ImageJ software. To categorize cell types expressing the WT (wild-type) and/or *LacZ* alleles in the brain tissue of heterozygous knockin-knockout mice, cells were classified according to five categories: (1) WT only: Cells only expressing the WT allele; (2) WT dominant cells: Cells dominantly expressing the WT allele with a least one *LacZ* mRNA staining dot; (3) Biallelic: Cells expressing relatively equal levels of the WT and *LacZ* alleles; (4) *LacZ* dominant cells: Cells dominantly expressing the *LacZ* allele with a least one WT mRNA dot; (5) *LacZ* only: Cells only expressing the *LacZ* allele. The number of cells for each class was summed from all of the images taken for a single animal. This summed number of cells per animal is presented in bar graph figures in Figures 3, 5, and S6.

Human Tissue Processing

Fresh frozen human brain tissue blocks from the DRN/vPAG (dorsal raphe nucleus/ ventral periaqueductal gray) region, BA (Broca's Area), OFC (orbitofrontal cortex) and ACC (anterior cingulate cortex) were embedded in OCT compound and cryosectioned at 14 microns thickness across a series of 12 slides, with two sections per slide. A 1 cm^2 region of tissue was traced out for each sample using a pap pen and stained using RNAscope in situ hybridization probes, as described above. The number of monoallelic versus biallelic cells for each gene analyzed was quantified per 1 cm^2 section for each individual and brain region.

Pyrosequencing

Allele-specific pyrosequencing was performed as previously detailed (Bonthuis et al., 2015). Briefly, we identified SNP sites that discriminate Cast and B6 alleles and amplified ~100-200 base pairs surrounding the SNP site. The sequencing primer was designed around the SNP site and the ratio of Cast versus B6 base calls was determined in each sample according to the Pyromark software. Amplification primers and sequencing primers are provided in Table S4. Pyrosequencing was performed and analyzed using the PyroMark Q24 system and software (QIAGEN) in the University of Utah DNA Sequencing Core Laboratory. We performed 5-8 biological replicates for each F1cb/F1bc cross for each gene and calculated the Cast:B6 allele expression ratio for each replicate. Using a ratiometric statistic for correlation (Abelin et al., 2014), we measured the dispersion of Cast:B6 allele expression ratios across the F1cb/F1bc replicates samples by computing the coefficient of variation (CV). The ratiometric statistic was computed as follows:

$$\Delta CV = \mathrm{abs}(CV(AE_P/AE_M) - CV(AE_M/AE_P)).$$

where

AE_P is the expression level of the paternal allele,
AE_M is the expression level of the maternal allele,
And abs is the absolute value.

The linear regression analysis comparing the RNA-seq r_a values to the PyroSEQ ΔCV values for the validated genes was performed using Prism software.

QUANTIFICATION AND STATISTICAL ANALYSIS

All n's in the study represent biological replicates involving different animals. Error bars represent standard error of the mean (SEM). Statistical analyses for the genomics are provided in detail above and appropriate sample sizes were estimated by statistical modeling using the procedures detailed above. RNA-seq experiments for F1cb and F1bc mouse samples were carried out with a randomized, blocked design on the flowcell to prevent batch effects causing systematic differences between F1cb and F1bc samples. Statistical analysis for cell counts involved a t test or a one-way ANOVA with a Tukey's HSD post-test, as indicated in the figure legends. Quantification was carried out in ImageJ (Schneider et al., 2012), as detailed in the Mouse and Human Tissue Processing subsections under METHOD DETAILS. To evaluate RNA quality in mouse and human tissue samples, an internal positive control probe that is provided with the ACDBio RNAScope kit was run for every staining reaction and samples that failed to show consistent

labeling across the tissue for the internal control probe were not included in the analysis. Blind quantifications were performed for cell counts comparing monoallelic and biallelic cells in brain tissue sections from animals at different ages (Figure 3). A power analysis for the cell quantification experiments was not performed.

DATA AND SOFTWARE AVAILABILITY

The BAM files for the adult mouse DRN, ARN, liver and muscle were published previously (Bonthuis et al., 2015). The data for the cerebellum was published previously (Perez et al., 2015). The P5 and P15 mouse DRN RNA-seq BAM files and macaque genome and RNA-seq BAM files have been deposited to the GEO database. The accession number for the data reported in this paper is GEO: GSE93788.

Neuron, Volume *93*

Supplemental Information

Diverse Non-genetic, Allele-Specific Expression Effects Shape Genetic Architecture at the Cellular Level in the Mammalian Brain

Wei-Chao Huang, Elliott Ferris, Tong Cheng, Cornelia Stacher Hörndli, Kelly Gleason, Carol Tamminga, Janice D. Wagner, Kenneth M. Boucher, Jan L. Christian, and Christopher Gregg

SUPPLEMENTAL DATA

Non-Genetic Allelic Effects Uncovered by RNASeq are Reproducible By Pyrosequencing

We tested whether AAEEs, DAEEs and allele CoEEs detected by RNASeq can be independently validated by pyrosequencing, an orthogonal approach. Pyrosequencing provides the relative ratio of maternal:paternal allele expression in F1cb and F1bc samples, which is not amenable to calculating an r_a value. Instead, we used a ratiometric statistic (ΔCV) based on the coefficient of variation (CV) (Abelin et al., 2014). For this approach, the Cast:B6 allele ratios were ascertained across F1cb/F1bc biological replicates using pyrosequencing (n=10-16 (5-8 per cross)) and the ΔCV statistic was calculated for each gene. In contrast to the Pearson r_a statistic, low ΔCV values indicate allele CoEEs and high values indicate DAEEs (Fig. S3B). For example, our adult female DRN RNASeq data indicates that *Faim2* exhibits robust allele CoEEs with an r_a = +0.95 (Fig. S3B). In contrast, the gene *Ide* has a low r_a value equal to -0.03, which is indicative of DAEEs (Fig. S3B). Using the ΔCV statistic to represent the same RNASeq data, we found that *Faim2* has a low ΔCV, while *Ide2* has a high ΔCV (Fig. S3C), consistent with CoEEs and DAEEs, respectively. Finally, we performed RT-PCR and pyrosequencing to analyze the allelic expression of both genes with an independent approach and with new RNA samples from different animals. The pyrosequencing results confirm the RNASeq results for both genes by showing that the ΔCV for *Faim2* is lower than the ΔCV for *Ide* (Fig. S3C). These results therefore indicate a methodology to independently validate the allelic effects detected by our RNASeq approach.

With this pyrosequencing and statistical approach, we performed independent validation studies of the allelic effects for 14 randomly selected, non-imprinted autosomal genes in adult female DRN and ARN (ARN samples indicated in Fig. S3D). The results show a significant negative linear relationship between the pyrosequencing ΔCV statistic and the RNASeq r_a statistic across all of these experiments (Fig. S3D). This result is expected if the allelic effects are reproducible with both approaches. Therefore, our results show that the allelic effects uncovered in our RNASeq screen are reproducible using an independent, orthogonal approach.

Developmental Changes to the Allelic Expression of X-linked Genes in the Female Brain

Due to the effects of random X-inactivation, we expect that most X-linked genes should exhibit AAEEs and have negative r_a values (Fig. 1A). Unexpectedly however, only 11% of X-linked genes meet the criteria for high confidence AAEEs in the adult female DRN. We investigated this result in more detail by examining the r_a values for X-linked genes at different ages and in different brain regions and tissues. Remarkably, we found that the distribution of r_a values for X-linked genes in the P5 DRN, and the adult female liver and muscle are shifted towards negative r_a values, as expected. In contrast, for the P15 DRN, adult DRN and ARN, we uncovered a significant shift towards more positive r_a values (Fig. S4A and C). These results therefore indicate that X-linked genes in the juvenile and adult female brain exhibit relatively increased allele CoEEs compared to X-linked genes in the neonatal brain, or in the adult liver and muscle. To determine whether this brain and developmental stage specific effect also occurs in the

cerebellum, we performed the same analysis for X-linked genes in the independent datasets from Perez and colleagues (Perez et al., 2015) and found the same developmental shift from AAEEs at P8 to increased allele co-expression in the adult cerebellum (Fig. S4B).

By tallying the number of X-linked genes with AAEEs, DAEEs and allele CoEEs at each age and in each tissue, we determined that AAEEs are prevalent in P5 DRN and in the adult liver and muscle; however very few AAEEs were detected in the P15 DRN and an increase in DAEEs emerged (Fig. S4C). This relative increase in the number of X-linked genes with DAEEs and reduction in AAEEs persists in the adult female DRN and ARN. In the ARN, a major increase in the number of uncategorized X-linked genes also emerges due to the shift toward increased allele co-expression (Fig. S4C). For all ages and tissues, very few X-linked genes exhibit allele CoEEs, one of the few examples is *Ddx3x,* which is known to escape X-inactivation in the mouse (Fig. S4D). Overall, the unexpected developmental shift in allelic effects that we observe could potentially involve genes escaping X-inactivation in subpopulations of cells in the brain at later developmental stages, which would cause an increase in allele co-expression at the tissue level. Currently however, the underlying mechanism is unclear.

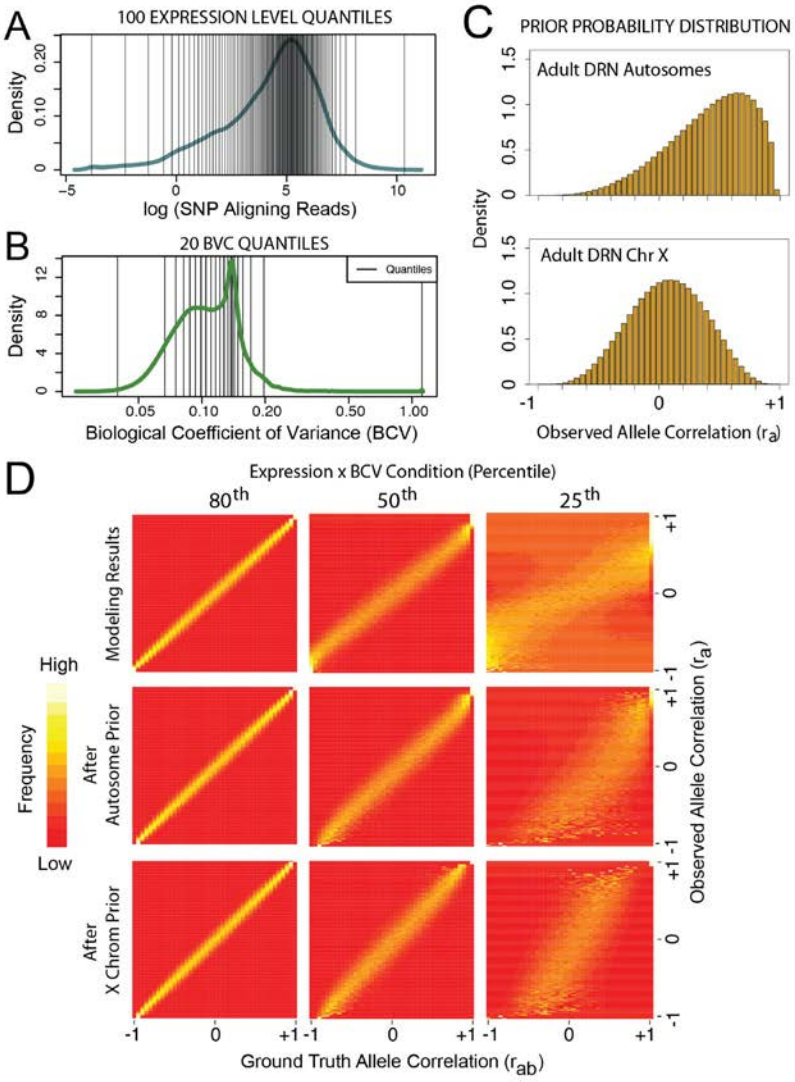

A
100 EXPRESSION LEVEL QUANTILES
Density
log (SNP Aligning Reads)
B
20 BVC QUANTILES
Quantiles
Density
Biological Coefficient of Variance (BCV)
C
PRIOR PROBABILITY DISTRIBUTION
Adult DRN Autosomes
Adult DRN Chr X
Density
Observed Allele Correlation (ra)
D
Expression x BCV Condition (Percentile)
80th
50th
25th
Modeling Results
After Autosome Prior
After X Chrom Prior
High
Frequency
Low
Observed Allele Correlation (ra)
Ground Truth Allele Correlation (rab)

FIGURE S1

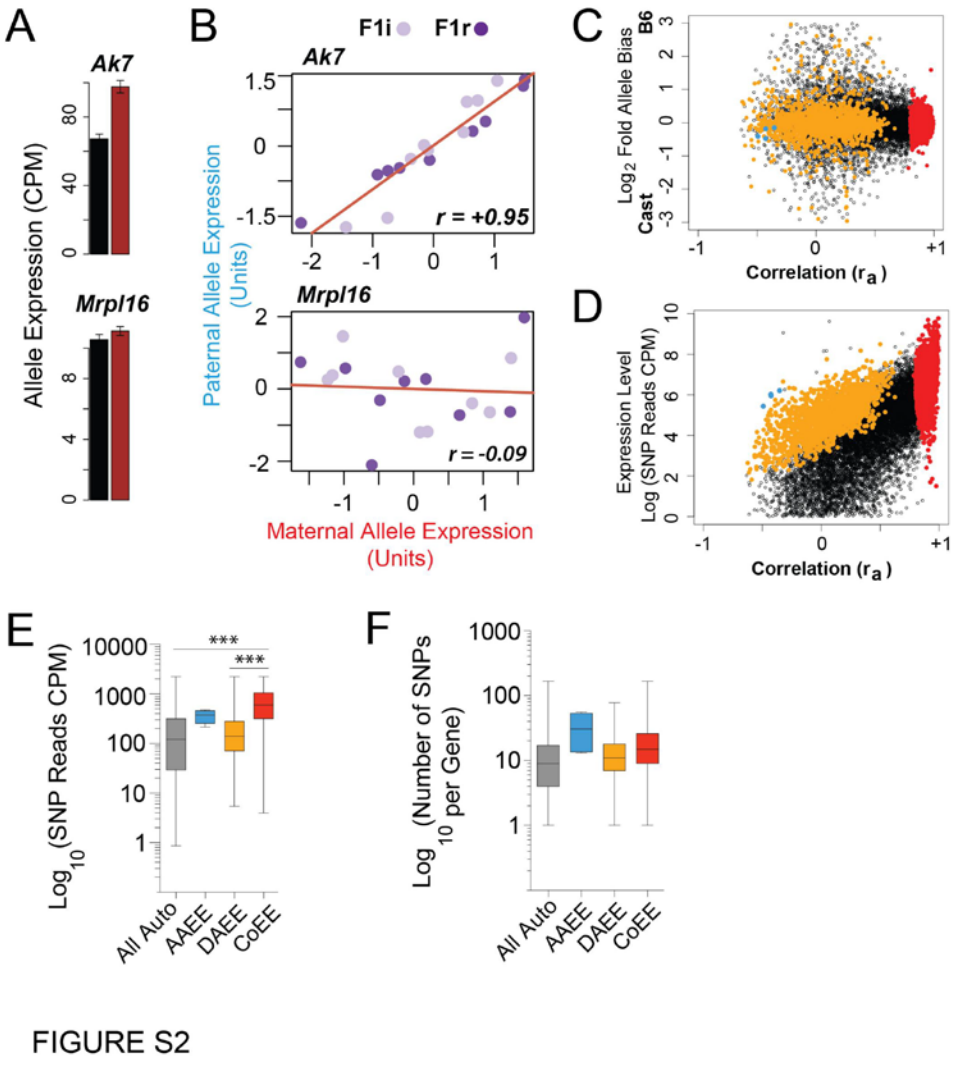

A
Ak7
Mrpl16
Allele Expression (CPM)
B
F1i
F1r
Ak7
r = +0.95
Mrpl16
r = -0.09
Paternal Allele Expression (Units)
Maternal Allele Expression (Units)
C
B6
Cast
Log2 Fold Allele Bias
Correlation (ra)
D
Expression Level
Log (SNP Reads CPM)
Correlation (ra)
E
Log10 (SNP Reads CPM)
***
***
All Auto
AAEE
DAEE
CoEE
F
Log10 (Number of SNPs per Gene)
All Auto
AAEE
DAEE
CoEE

FIGURE S2

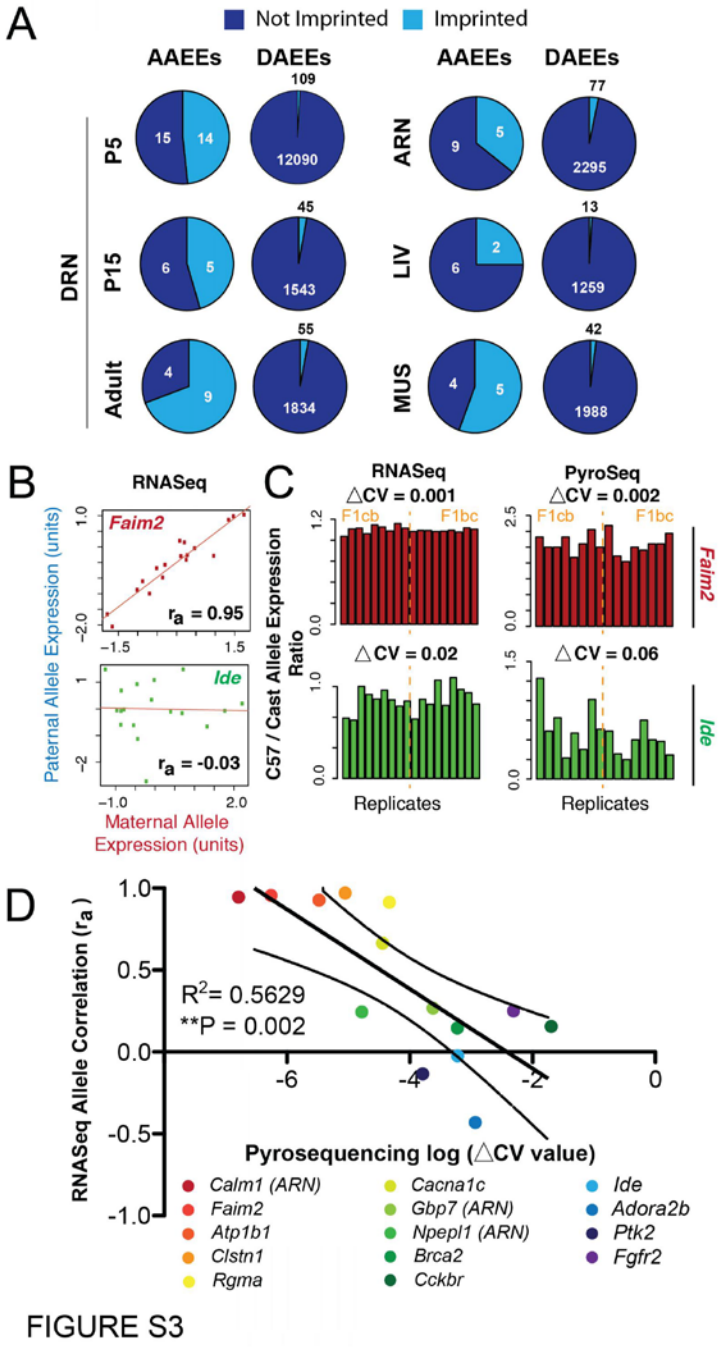

A
Not Imprinted
Imprinted
AAEEs
DAEEs
AAEEs
DAEEs
DRN
P5
P15
Adult
ARN
LIV
MUS
15
14
109
12090
9
5
77
2295
6
5
45
1543
6
2
13
1259
4
9
55
1834
4
5
42
1988
B
RNASeq
Faim2
r_a = 0.95
Ide
r_a = -0.03
Paternal Allele Expression (units)
Maternal Allele Expression (units)
C
RNASeq
△CV = 0.001
PyroSeq
△CV = 0.002
F1cb
F1bc
△CV = 0.02
△CV = 0.06
C57 / Cast Allele Expression Ratio
Replicates
Faim2
Ide
D
RNASeq Allele Correlation (r_a)
R^2= 0.5629
**P = 0.002
Pyrosequencing log (△CV value)
Calm1 (ARN)
Faim2
Atp1b1
Clstn1
Rgma
Cacna1c
Gbp7 (ARN)
Npepl1 (ARN)
Brca2
Cckbr
Ide
Adora2b
Ptk2
Fgfr2

FIGURE S3

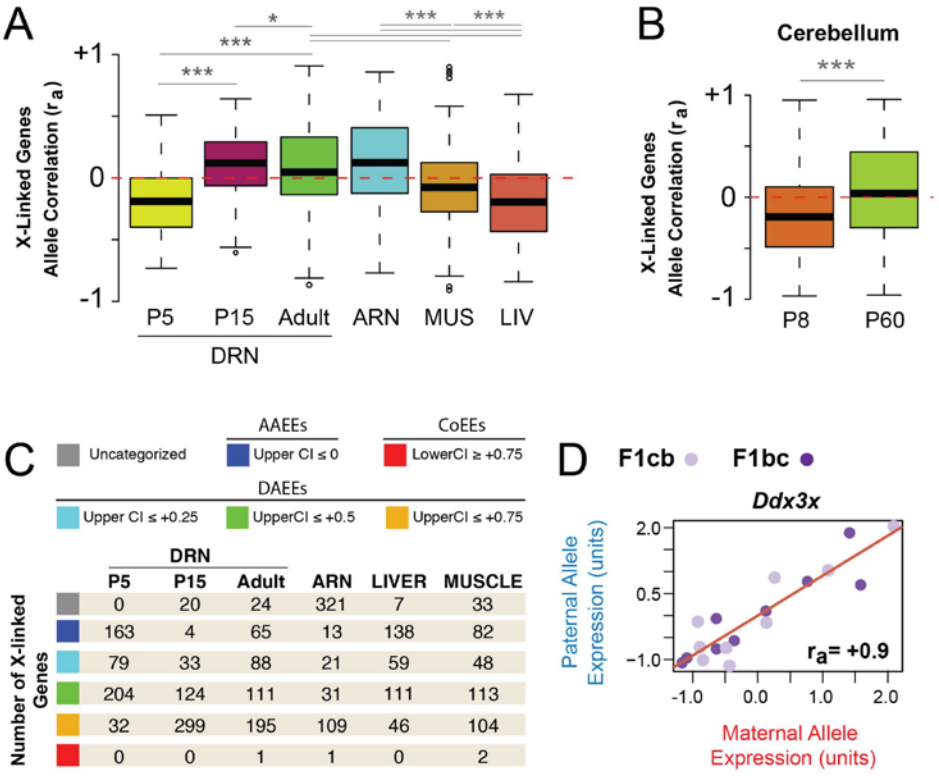

A
X-Linked Genes
Allele Correlation (r a)
+1
0
-1
*
***
***
***
***
P5
P15
Adult
ARN
MUS
LIV
DRN
B
Cerebellum
***
P8
P60
C
AAEEs
CoEEs
Uncategorized
Upper CI ≤ 0
LowerCI ≥ +0.75
DAEEs
Upper CI ≤ +0.25
UpperCI ≤ +0.5
UpperCI ≤ +0.75
DRN
P5 P15 Adult ARN LIVER MUSCLE
Number of X-linked Genes
0 20 24 321 7 33
163 4 65 13 138 82
79 33 88 21 59 48
204 124 111 31 111 113
32 299 195 109 46 104
0 0 1 1 0 2
D
F1cb
F1bc
Ddx3x
Paternal Allele Expression (units)
2.0
0.5
-1.0
-1.0
0.0
1.0
2.0
ra= +0.9
Maternal Allele Expression (units)

FIGURE S4

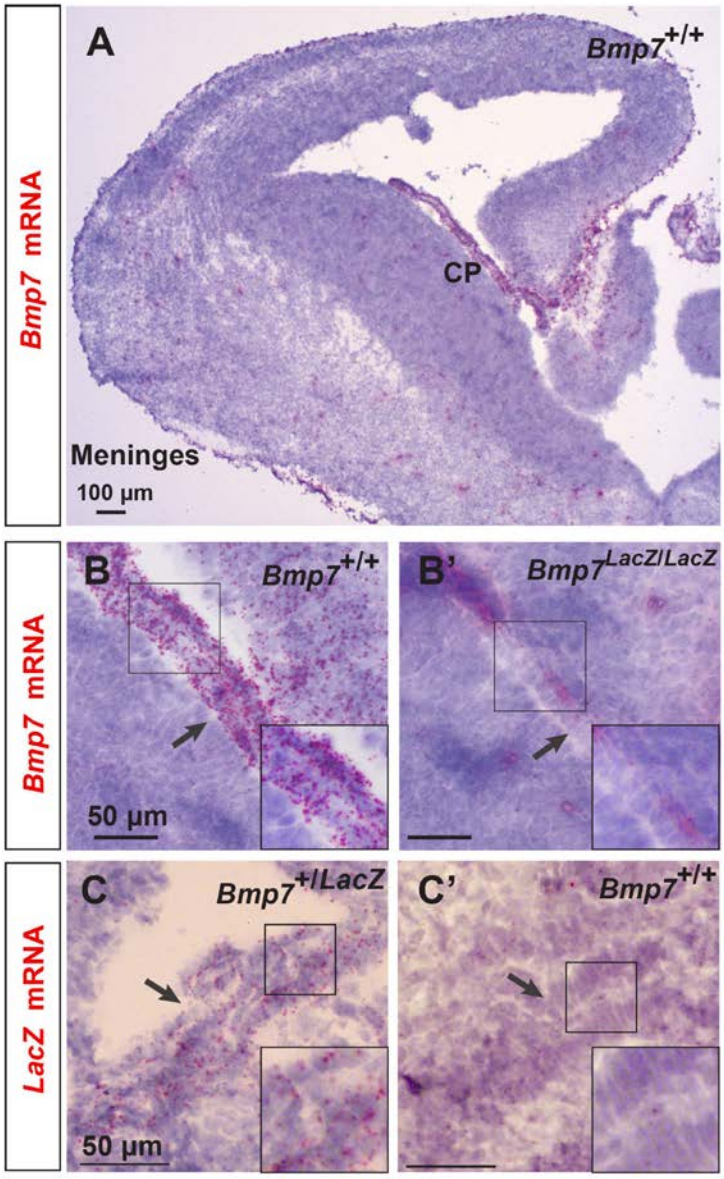

Bmp7 mRNA
A
Bmp7+/+
CP
Meninges
100 µm
Bmp7 mRNA
B
Bmp7+/+
50 µm
B'
Bmp7LacZ/LacZ
LacZ mRNA
C
Bmp7+/LacZ
50 µm
C'
Bmp7+/+

FIGURE S5

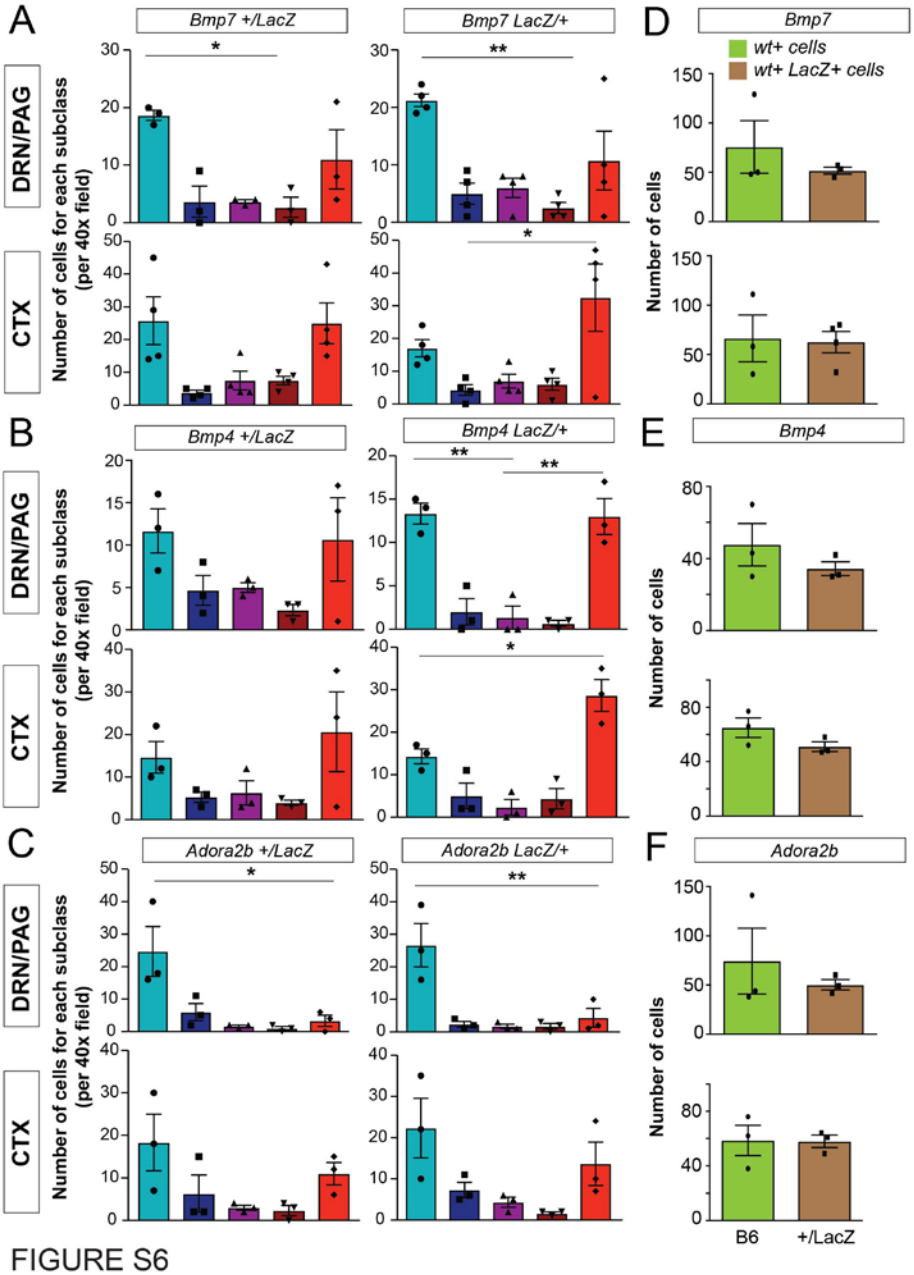

A
Bmp7 +/LacZ
Bmp7 LacZ/+
D
Bmp7
wt+ cells
wt+ LacZ+ cells
DRN/PAG
CTX
Number of cells for each subclass (per 40x field)
Number of cells
B
Bmp4 +/LacZ
Bmp4 LacZ/+
E
Bmp4
C
Adora2b +/LacZ
Adora2b LacZ/+
F
Adora2b
B6
+/LacZ

FIGURE S6

SUPPLEMENTARY FIGURE LEGENDS

Figure S1. An empirical Bayesian approach to define the 95% CI for the ground truth biological non-genetic allelic effects for each gene. Related to Figure 1 and STAR methods.

(A) Density plot of the distribution of mean gene expression levels based on SNP-aligning RNASeq reads. The data depicted is adult female F1cb and F1bc DRN RNASeq data for ~15,000 genes, which is divided into 100 quantiles of genes according to different expression levels. The quantiles of different mean expression levels are used for modeling the effects of biological and technical variance on genes with different expression level characteristics.

(B) Density plot of the distribution of the biological coefficient of variance (BCV) for each gene expressed in the adult female F1cb and F1bc DRN. The data is divided into 20 different quantiles that are used to model the effects of biological and technical variance on genes with different variance characteristics.

(C) The distributions of r_a values from adult female DRN data for autosomal and X-linked genes. These distributions were used as priors to weight the frequency distribution results from the modeling.

(D) Representative tables of the modeling results for genes with different expression level x BCV characteristics in the RNASeq data. Examples are shown for genes in the 80^{th}, 50^{th}, and 25^{th} expression level x BCV percentiles. The heatmaps (scaled by row maximum) for each table show the distribution frequency of the observed r_a values (y-axis) for all possible ground truth allele correlation effects (r_{ab}; x-axis). The results from the gamma-Poisson mixed modeling approach used to capture biological and technical

variance are depicted (top row). In addition, the results are shown after the application of the empirical prior probability distribution for autosomal genes (middle row) and for X-linked genes (bottom row). Only the bottom 25th percentile of genes are substantially impacted by the prior.

Figure S2. Genes with high confidence AAEEs or DAEEs are not enriched for genes with strain-based genetic allelic effects, low expressed genes or genes with few SNPs. Related to Figure 2.

(A and B) Genes with significant differences in the expression of Cast (brown bar) versus B6 (black bar) alleles in F1 hybrid mice, such as *Ak7,* can exhibit allele CoEEs, while genes with no genetic strain effects on allele expression can exhibit DAEEs, such as *Mrpl16.* This result is expected since F1cb and F1bc mice are genetically identical and therefore the genetic effects on Cast versus B6 allele expression are the same in each individual mouse.

(C) Scatterplot of the fold bias in Cast versus B6 allele expression versus the r_a values for all autosomal genes (black dots) in the adult female DRN. Non-imprinted genes with high confidence AAEEs (blue), DAEEs (orange) and CoEEs (red) are indicated. The results show that genes with AAEEs or DAEEs are not enriched for genes with strong Cast versus B6 allele biases and most have very modest strain-based allele biases (Allele bias near 0).

(D and E) Scatterplot of gene expression level (based on SNP-aligning reads) versus r_a values for all autosomal genes (black dots) in the adult female DRN (D). Genes with AAEEs or DAEEs are not enriched among genes with low expression levels (D). A

boxplot comparison of the expression levels for all autosomal genes (grey) versus genes with AAEEs, DAEEs and CoEEs reveals significantly increased expression levels for genes with CoEEs (E). One-way ANOVA with Tukey HSD. ***P < 0.001. CPM, counts per million.

(F) A boxplot comparison of the number of SNPs per gene for all autosomal genes (grey) versus genes with AAEEs, DAEEs and CoEEs reveals that genes with AAEEs or DAEEs are not enriched for genes with few SNPs. Thus, genes with ASE effects are not characterized by potentially reduced fidelity of allelic expression detection by RNASeq (Median number of SNPs per gene = All: 9, AAEEs: 31, DAEEs: 11, CoEEs: 15).

Figure S3. Most autosomal DAEEs are not due to genomic imprinting, and are reproducible by pyrosequencing. Related to Figure 2 and supplemental results.

(A) Pie charts indicating the number of autosomal genes with AAEEs or DAEEs that are imprinted, and the number that are not imprinted, for the female P5 DRN, P15 DRN, and adult DRN, as well as the adult ARN, liver and muscle.

(B and C) Examples of genes with allele CoEEs (*Faim2*) or DAEEs (*Ide*) detected by RNASeq in the adult female DRN (B). Illustration of an alternative approach to allele co-expression analysis by evaluating the Cast:B6 allele expression ratio across F1cb and F1bc replicates and using the ratiometric statistic (ΔCV) to examine differences in the variance of the allele expression ratio across replicates (C). Genes with CoEEs (*Faim2*) have a low ΔCV and genes with DAEEs (*Ide*) have a relatively high ΔCV (C). This approach allows us to analyze allele co-expression using pyrosequencing in F1cb and F1bc mice (PyroSeq). PyroSeq analysis for *Faim2* and *Ide* reveals that allelic effects

detected by RNASeq are reproducible using an independent approach. (N=8 F1cb and N=8 F1bc replicates per PyroSeq experiment).

(D) Allele CoEEs and DAEEs detected by RNASeq are reproducible by PyroSeq. Fourteen randomly selected, non-imprinted autosomal genes with CoEEs or DAEEs detected by RNASeq were analyzed by PyroSeq for the adult female DRN or ARN (N = 10-16 (5-8 per cross)). The ΔCV was computed for each gene (see C and Supplemental Information). A plot of the RNASeq r_a values versus the PyroSeq ΔCV values for each gene reveals a statistically significant negative linear relationship, indicating that the RNASeq results are reproducible by PyroSeq.

Figure S4. Identification of increased allele co-expression for X-linked genes in the female juvenile and adult brain compare to the neonatal brain and adult liver and muscle. Related to Figure 2 and supplemental results.

(A) Boxplots of the r_a values for all X-linked genes expressed for each age and tissue reveal a negatively shifted distribution of r_a values for the P5 DRN and for the adult muscle and liver, consistent with random X-inactivation effects. However, for the P15 DRN, adult DRN and adult ARN, the r_a distribution is shifted toward increased allele co-expression.

(B) Boxplots of the r_a values for all X-linked genes expressed in the P8 versus adult (60) female cerebellum. The data independently confirms a developmental shift from negative r_a values for X-linked genes in P8 pups to increased allele co-expression for X-linked genes in P60 adults, as indicated by higher r_a values.

(C) The number of X-linked genes in each high confidence allelic expression category, according to age and tissue type, reveals brain and developmental stage specific changes in allelic expression for X-linked genes.

(D) Example of an X-linked gene with high confidence CoEEs in the adult female DRN. *Ddx3x* is known to escape X-inactivation in the mouse and the maternal and paternal alleles for *Ddx3x* exhibit correlated expression.

Figure S5. Semi-quantitative RNAscope in situ hybridization probes targeting mRNA are specific and have robust signal:noise detection properties. Related to Figure 4.

(A and B) RNAscope probe labeling for *Bmp7* in the embryonic day(E)14.5 brain of wild-type embryos. Robust staining is observed in the choroid plexus (CP) and meninges, which are known to express high levels of *Bmp7* (A). Specific staining is observed in the CP of wild-type embryos (B) and absent in the CP of E14.5 $Bmp^{LacZ/LacZ}$ embryos (B').

(C and C') RNAscope probe labeling for *LacZ* mRNA in the CP of E14.5 $Bmp7^{+/LacZ}$ heterozygotes (C) and E14.5 $Bmp^{+/+}$ wild-type embryos (C'). Robustly labeled cells in the CP are evident in $Bmp7^{+/LacZ}$ embryos and easily identifiable by multiple staining dots (transcripts) in each positive cell (C). Very little non-specific background labeling is observed in E14.5 wild-type CP, and for the non-specific background that is present the labeling appears as single dots evenly distributed across the tissue, presumably due to small amounts of remaining probe following washing (C').

Figure S6. Allelic effects in *Bmp7, Bmp4* and *Adora2b* knockout-reporter mice are not related to genomic imprinting or ectopic expression of the *LacZ* transgene. Related to Figure 5.

(A-C) Number of monoallelic-WT, WT dominant, biallelic, *LacZ* dominant and monoallelic-*LacZ* expressing cells in reciprocal heterozygote knockout-reporter mice with a maternally (*LacZ*/+) versus paternally (+/*LacZ*) inherited mutant *LacZ* allele. Cells in each category were quantified in the DRN/PAG and CTX for *Bmp7* (A), *Bmp4* (B) and *Adora2b* (C) reciprocal heterozygous knockout-reporter mice. Parent-of-origin effects that influence the relative proportion of the cells categorized by allelic expression are not observed, and monoallelic-WT and monoallelic-*LacZ* allele expressing cells are observed in both crosses for all three transgenic lines. (N= 3-4, one-way ANOVA and Tukey HSD post test). See Legend in Figure 5A. *P<0.05, **P<0.01

(D-F) Barplots showing the number of cells expressing the WT allele (*Bmp7, Bmp4* or *Adora2b*) in the DRN/PAG of adult female B6 control mice (green bar) compared to the number of cells expressing the WT allele and/or the *LacZ* allele in adult female +/LacZ knockout-reporter mouse lines (brown bar) for *Bmp7* (D), *Bmp4* (E), and *Adora2b* (F). No significant difference in the number positive cells was observed, indicating that the *LacZ* mutant allele is not labeling an additional ectopic cell population. (N=3-4, two tailed t-test).

SUPPLEMENTAL TABLE LEGENDS

Supplementary Table S1. Maternal and paternal allele co-expression data for the mouse. Related to Figures 1 and 2. This table presents the results of our allele co-

expression analysis for every gene expressed for the female mouse P5, P15 and adult DRN and for the adult female ARN, liver and muscle. Column definitions: EnsembleID; MeanExpressionLevel_SNPAligningReads_CPM, counts per million for SNP aligning reads presented as the mean across all replicates; BCV, biological coefficient of variation calculated using edgeR across all replicates; ra_value, Pearson correlation for the observed maternal and paternal allele co-expression; Ratiometric_deltaCV_value, ratiometric statistic for maternal and paternal allele co-expression; external_gene_id, gene name; Chr, chromosome; rab_value_95%CI_LowerLimit, the lower limit of the r_{ab} value 95% confidence interval computed from modeling; rab_value_95%CI_UpperLimit, the upper limit of the r_{ab} value 95% confidence interval computed from modeling; CI_Width, the width of r_{ab} value 95% confidence interval; ImprintedFDR05, true/false column indicating significantly imprinted genes; X_Auto, indicates autosomal or X-linked gene; Age, indicates age of mouse (P5, P15 or adult); Tissue, indicates tissue or brain region type; CoEEs, true/false indicating high confidence allele co-expression effects; DAEEs, true/false indicating high confidence differential allele expression effects; AAEEs, true/false indicating high confidence antagonistic allele expression effects.

Supplementary Table S2. Maternal and paternal allele co-expression data for the cynomolgus macaque. Related to Figures 6 and 7. This table presents the data for the allele co-expression analysis for genes expressed in the juvenile female macaque dorsal raphe nucleus (DRN) region of the brain. Only genes with allele discriminating SNPs in all ten of the daughters analyzed are included. Column definitions: GeneName; Log2_MeanExpressionLevel_SNPAligningReads_CPM, counts per million for SNP

aligning reads presented as log2 of the mean across all replicates; BCV, biological coefficient of variation calculated using edgeR across all replicates; ra_value, Pearson correlation for the observed maternal and paternal allele co-expression; Ratiometric_deltaCV_value, ratiometric statistic for maternal and paternal allele co-expression; external_gene_id, gene name; Chr, chromosome; rab_value_95%CI_LowerLimit, the lower limit of the r_{ab} value 95% confidence interval computed from modeling; rab_value_95%CI_UpperLimit, the upper limit of the r_{ab} value 95% confidence interval computed from modeling; CI_Width, the width of r_{ab} value 95% confidence interval; X_Auto, indicates autosomal or X-linked gene; Age, indicates age of mouse (P5, P15 or adult); Tissue, indicates tissue or brain region type; CoEEs, true/false indicating high confidence allele co-expression effects; putative DAEEs, true/false indicating high confidence putative differential allele expression effects; DAEEs, true/false indicating high confidence differential allele expression effects; AAEEs, true/false indicating high confidence antagonistic allele expression effects. Category, column indicating the allelic category for each gene.

Supplementary Table S3. Analysis of in vivo r_a values in the mouse or primate DRN for published random monoallelic genes identified in mouse or human cell lines. Related to data in Figures 2 and 6. This table presents the results of an analysis of the female adult mouse or juvenile primate DRN r_a values for autosomal genes previously identified as random monoallelic in published studies of mouse or human cell lines (see Discussion for details). The r_a values for published random monoallelic genes are compared to the r_a values for all expressed autosomal genes to determine whether

the published random monoallelic genes have significantly lower r_a values, as would be expected if they differ in their allelic expression compared to other autosomal genes in the tissue analyzed. The citations for the papers from which the gene lists were obtained are indicated.

Supplementary Table S4. Descriptive statistics for genomics experiments and pyrosequencing primers. Related to Figures 1, S1, S2, 2, S3 and 6. Includes number of uniquely aligned reads for each mouse RNASeq replicate and metadata for macaque parent and daughter genomics study. Mouse RNASeq, poly(A)-selected RNASeq 50 bp single-end reads; Primate whole genome, 100 bp paired-end reads; Primate RiboZero-treated whole transcriptome RNASeq, 100 bp paired-end reads.

Supplementary Table S5. Estimated proportion of genes with genetic effects impacting allelic expression in each cynomolgus macaque daughter. Related to Figure 6 and the STAR methods. This table presents the estimated variant rate per kilobase in macaque promoter regions (VR_{reg}) for each daughter and the estimated probability that a given gene has a genetic allelic effect in each daughter (see methods for details).

Supplementary Table S6. Cast versus B6 Genetic Allele Expression Effects in the Adult Female Mouse DRN. Related to Figure 6 and the STAR methods. This table presents the magnitude of the Cast versus B6 genetic allelic effect for each gene expressed in the adult female mouse DRN region of the brain. These genetic allelic

effects in the mouse are used to model genetic allelic effects in the primate, as detailed in the STAR methods section. The data is calculated from the study by Bonthuis et al. (2015). Column definitions: Gene, indicates ensemble gene IDs; log2 Fold Allele Expression Bias, positive values indicate a Cast allele bias and negative values indicate a B6 allele bias; logCPM, is the log 10 counts per million expression level estimation based on SNP-aligning reads; PValue, is the edgeR generalized linear model p-value for a main effect of strain on allelic expression; FDR, Benjamini-Hochberg false discovery rate.

CHAPTER 4

DISCUSSION

Part of this chapter will be incorporated in a review in preparation Gregg C & Huang WC *Trends in Neurosciences*, 2018.

4.1 Summary

Epigenetic ASEs, such as genomic imprinting and RME, represent a noncanonical biological phenomenon in which two homologous genes located in the same cell exhibit distinctive expression patterns in either a deterministic or random manner. The genes subject to ASEs have a variety of functions that influence a wide range of phenotypes and complex disorders. However, the repertoire and temporal and spatial variability of ASE genes in vivo have not been characterized fully. The roles ASE genes play in cell physiology and the way in which ASE gene mutations affect genetic architecture remain unclear as well. Therefore, this dissertation work developed a statistical framework to identify putative RME genes in vivo genome-wide and revealed several novel ASE genes in different tissues and different developmental stages (Chapter 3). To validate the method's reliability in identifying RME genes, pyrosequencing was used as an independent approach to confirm the RME profiles of several of the genes identified (Chapters 2 & 3). Moreover, different mouse strains (wild mice, Chapter 2) and different species (*Macaca cynomolgus*, Chapter 3) were examined to confirm that the framework can be generalized to multiple contexts. To determine the nature of ASE genes at the cellular level, a highly sensitive in-situ hybridization method with cellular resolution was performed on tissue sections to observe allelic expression in subpopulations of cells in vivo (Chapters 2 & 3). Finally, heterozygous transgenic mice were used to determine the way in which mutation of ASE genes influences behavioral phenotypes (Chapter 2, performed by Dr. Paul Bonthuis) and genetic architecture (Chapter 3). The findings of this dissertation expand our understanding of the spectrum of ASE genes in vivo and the potential roles ASE genes play in function.

4.2 Monoallelic expression at the cellular level

In the two studies performed, RNA in-situ hybridization was used to demonstrate that ASE genes, such as imprinted and RME genes, express one allele in a subpopulation of brain cells preferentially. This evidence suggests that noncanonical imprinted genes that are found at the tissue level and exhibit expression differences between the two parental-inherited alleles, described as parental-biased genes in the introduction, may result from monoallelic expression within a particular subpopulation of cells, while other cells display biallelic expression. Similarly, the proportion of cells that express monoallelically corresponds to the coexpression correlation between two alleles in the tissue level, suggesting that ASE genes do not necessarily exhibit absolute monoallelic expression at the cellular level. Based on these findings, the next intriguing question is how can we identify the cell lineages that express monoallelically or biallelically preferentially? Further, do any cell lineages exist that express the maternal or paternal allele preferentially to lead to cell type-specific genomic imprinting? This scenario also could occur in RME genes that may result from random pooling of several cell lineages that express an allele preferentially following the parent-of-origin pattern.

4.2.1 Allele-specific in-situ hybridization

Despite the fact that our novel ultra-sensitive in-situ hybridization provides resolution at the cellular level to observe allele expression in an individual cell, the approach cannot obtain the parental information of the transcript expressed. One imaging approach to discriminate among transcripts with small sequence differences, particularly those of low abundance, is single-molecule allele-specific in-situ hybridization. The first

single-molecule allele-specific in-situ hybridization that used fluorescent probes was developed approximately ten years ago (Bonifazi et al., 2006); however, the resolution of images obtained at that time still needs to be improved. Recently, several approaches have been proposed to strengthen the signal from the probes targeting allelic differences. Nilsson and colleagues performed a padlock-probe in-situ to detect a somatic point mutation in mouse and human cells (Larsson, Grundberg, Soderberg, & Nilsson, 2010), while Levesque et al. invented the toehold probe strategy to visualize allele expression in a human lymphoblast cell line (Levesque, Ginart, Wei, & Raj, 2013) and used the same approach recently to detect loss of imprinting of *H19* in cells harboring mutations in the CCCTC-binding factor (CTCF) sites in the *H19/IGF2* imprinting locus (Ginart et al., 2016). Hansen and van Oudenaarden used standard DNA probes to observe maternal and paternal alleles of the Nanog gene in mouse embryonic stem cells and to demonstrate the expression bias of two alleles under different culture conditions (Hansen & van Oudenaarden, 2013). One limitation of allele-specific in-situ hybridization is that the method primarily provides a snapshot from fixed or frozen tissues rather than continuous recording of the same cell to observe dynamics of allele expression. Even so, multicolor, allele-specific in-situ hybridization provides a powerful tool to observe mosaicism of allele expression at the tissue level.

4.2.2 Single cell sequencing

As mentioned in the introduction, single cell RNA sequencing provides high resolution of intercellular heterogeneity and a transcriptome profile that can estimate genome-wide allele expression with parental information. Single-cell RNA sequencing

has been used in several ASE studies, such as RME (Deng, Ramskold, Reinius, & Sandberg, 2014), genomic imprinting (Santoni et al., 2017), and X-inactivation (Petropoulos et al., 2016). Although single cell RNA sequencing can obtain genome-wide information at a specific time, it lacks temporal information to observe the change of expression in the cell. Deng et al. performed single cell RNA-seq on blastomeres and found that allelic expression patterns varied across the cells divided from the same embryo (Deng et al., 2014). This finding differed from the results of a cell culture, which showed that the allelic expression pattern is identical in all cells derived from a single cell. Therefore, the authors proposed a novel mechanism referred to as dynamic RME, which is caused by asynchrony of burst frequency and discrepancy of the RNA degradation rate between two alleles (Reinius & Sandberg, 2015). However, the evidence from single cell RNA-seq cannot exclude the possibility that each cell owns a unique allelic expression barcode so that the identical expression pattern is found rarely. Thus, time-lapse analysis of imaging is still required to prove the existence of dynamic RME.

Allelic expression heterogeneity was observed in several single-cell transcriptome studies, although the epigenetic mechanisms underlying the extent of this heterogeneity remain unclear. Integration of several sequencing data collected from single cells has enabled us to link individual allele expression pattern to the distinct epigenetic signature. Angermueller et al. developed the scM&T-seq, which involves parallel, single-cell transcriptome sequencing and genome-wide methylome to observe the link between transcriptional and epigenetic variation in mouse embryonic stem cells (Angermueller et al., 2016). In a more recent study, Guo et al. developed multiomics sequencing to perform single-cell analysis of genome-scale chromatin accessibility and DNA

methylation dynamics (Guo et al., 2017). These methods have enabled the study of allele expression and epigenetic markers from the same cell.

4.2.3 Reporter line

In addition to allele-specific hybridization, reporter mouse lines allow single-cell visualization of ASE and isolation of cells by sorting. Recently, several reporter mouse lines have been established to study ASE. An X-linked reporter system in combination with cell type-specific Cre drivers was generated to analyze the mosaicism of X chromosome inactivation (XCI) in different tissues and cell types (Wu et al., 2014). Swanzey and Stadtfeld developed a reporter mouse model to visualize imprinted gene *Dlk1* by inserting fluorescent reporter genes into the exon 6 of *Dlk1* (Swanzey & Stadtfeld, 2016), while Jaenisch and colleagues used CRISPR/Cas9-mediated gene editing to establish a DNA methylation reporter and uncovered tissue- and cell-dependent allele expression during different developmental stages (Stelzer et al., 2016). In a recent study, luciferase was knocked into imprinted gene *Cdkn1c* locus to generate Cdkn1c-luciferase mice that offers noninvasive readouts at whole-body and single-cell resolution of CDKN1C (Van de Pette et al., 2017). Although reporter mouse lines allow longitudinal imaging and cell isolation, generation of reporter lines would be time-consuming.

4.3 Functional significance of RME

4.3.1 Diversity at the tissue, specificity at the cell

RME's functional significance has been implicated in known RME genes, immunoglobulins, clustered protocadherin, and olfactory receptors. The "one immune cell-one antigen receptor," or "one neuron-one olfactory receptor" principles guarantee cell specificity and prevent cells from wasting resources and misinterpreting information. Similarly, protocadherins function as a unique address for individual neurons in neuronal connection formations during development. On the other hand, these three groups of RME genes maximize the variability of allele expression to maintain diversity at the tissue level. Thus, the olfactory system enables us to discern differences among mixtures of odors in the real world, and the adaptive immune system can respond to most pathogens that invade the body. One shared characteristic of these three groups is stochastic expression. However, if stochastic choice occurred in each allele, at least two gene products would be expressed in the same cell and monospecificity would be lost. Therefore, RME serves as a selection mechanism to ensure that only one gene product is expressed in the cell. This working model can be used to explain other RME genes scattered in the genome. Gene isoforms are forms of messenger RNA (mRNA) spliced alternatively that are produced from the same transcriptional unit, but differ in protein coding DNA sequences or un-translated regions because of the alternative splicing (Lee & Rio, 2015), which is a typical scenario in eukaryotes that increases biodiversity. With the advance of next generation sequencing, more than 100,000 isoforms have been identified in humans, and > 90% human genes are found to express at least one isoform (Pan, Shai, Lee, Frey, & Blencowe, 2008; Wang et al., 2008). Many alternative isoforms

have been reported to yield products with novel protein sequences and have physiological functions (Kriventseva et al., 2003). Interestingly, increasing evidence has shown that most alternative splicing is a result of stochastic binding in the splicing machinery to generate alternative isoforms (Melamud & Moult, 2009; Pickrell, Pai, Gilad, & Pritchard, 2010). In this case, RME could play a role in selecting only one isoform expression in the cell for those genes that require monospecificity of isoform expression.

4.3.2 Heterogeneity

One intuitive thought about the functional consequence of RME, but not genomic imprinting, is that RME may lead to phenotypic heterogeneity in haploinsufficiency disorders, whereby a diploid organism shows a phenotype when one copy of a gene is lost. As described previously, a diploid system has two copies of most genes, and the healthy backup copy can reduce the effect of heterozygous mutations. However, if the healthy allele is silenced and the mutated allele is the only one expressed, then a heterozygous mutation could have a stronger effect. On the other hand, if the healthy allele is activated, but the mutated allele is silenced, the phenotype alteration would be relatively minor. One remarkable example is based on random XCI in females. In women with a heterozygous mutation of X-linked tumor suppressor gene *FOXP3*, breast cancer grows largely from cells with one silenced functional copy and one activated mutant copy, which results in a null mutation-like condition (Zuo et al., 2007). Thus, the severity of the phenotype in the organism would depend on the stochastic choice that determines the proportion of cells that inherited a functional or mutated copy (Medema & Burgering, 2007). The more cells with the mutated allele, the more severe the organism's phenotype.

4.4 Dynamics of allelic expression

4.4.1 Development

Developmental control of genomic imprinting and XCI has been studied well in the prenatal period. Genomic imprinting is erased in primordial germ cells and then reestablished during gametogenesis. As described above, there are two waves of XCI during embryonic development. The first occurs during the 4–8-cell stage and results in imprinted paternal XCI. The silenced paternal X chromosome is reactivated at the blastocyst stage, after which the second wave results in random XCI of either the maternal or paternal X chromosome. Some autosomal genes have been observed to change ASE during differentiation or development. NANOG, a stem cell pluripotent factor, showed monoallelic expression in embryonic stem cells cultured with medium containing serum and leukemia inhibitory factor (LIF), and then switched to biallelic expression in conditions with LIF and 2i (pharmacological inhibition of GSK-3 and MAPK; Miyanari & Torres-Padilla, 2012), although this result is still under debate (Faddah et al., 2013; Filipczyk et al., 2013; Hansen & van Oudenaarden, 2013). Transcription factor GATA3 was found to be expressed monoallelically in early T-cell progenitors**,** but switched to biallelic expression at midthymopoiesis (Ku et al., 2015). The dynamics of ASE can be observed transcriptome-wide with RNA-seq. Two recent studies used RNA-seq and estimated that ~0.5-1% of genes in embryonic stem cells and ~5–10% of genes in neural progenitor cells are subject to RME (Eckersley-Maslin et al., 2014; Gendrel et al., 2014), indicating that ASE changes widely when embryonic stem cells differentiate into neural progenitor cells. Our study found that RME genes are more prevalent in postnatal day (P) 5 brains compared to P15 and adult brains, suggesting that

RME is regulated developmentally in vivo. These temporal dynamics were observed with genomic imprinting in the postnatal brain as well. Perez et al. performed RNA-seq and profiled imprinted genes in the developing (P8) and adult (P60) mouse cerebellum to obtain a picture of developmental changes in genomic imprinting in the brain (Perez et al., 2015). Surprisingly, the comparison revealed that approximately half of all the imprinted genes identified in the cerebellum are regulated developmentally between P8 and P60. Taken together, these results highlight the dynamic nature of ASE during development.

4.4.2 Disorders

Disturbance of the normal balanced condition with respect to maternal and paternal allelic copies in the genome can cause deleterious outcomes, for example, nonviable gynogenetic or androgenetic embryos, as noted in the introduction. Studies of human imprinting disorders have shown that the disturbance of imprinted gene expression often leads to morbidity. Beckwith-Wiedemann syndrome (BWS) is an imprinting disorder caused by dysregulation at the *IGF2/H19* imprinted locus, and is characterized by overgrowth at birth and an increased risk of childhood cancer. One of the most common epigenetic disturbances in BWS patients is the loss of insulin growth factor 2 (*IGF2*) imprinting, which results in activation of the normally silenced maternal allele and overexpression of *IGF2* (Steenman et al., 1994). *IGF2* is a growth factor that promotes the growth of various cell types, and its overexpression is associated with several adult and childhood cancers. Indeed, loss of *IGF2* imprinting has been defined as a potential biomarker of increased risk for several cancers (Cui, 2007), such as colorectal

cancer (Cui et al., 2003), bladder cancer (Byun et al., 2007), and acute lymphoblastic leukemia (Vorwerk et al., 2003). This phenomenon suggests that loss of imprinting could be a predictive tool in detecting epigenetic instability within cancers. In addition to *IGF2*, loss of imprinting at the *MEG3-DLK1* imprinted locus was discovered in hepatocellular carcinoma (Anwar et al., 2012) and acute promyelocytic leukemia samples (Manodoro et al., 2014), while loss of imprinting at the *PEG1/MEST* locus also has been found in breast and lung cancer tissues (Kohda et al., 2001; Pedersen et al., 1999). Although the disturbance of several imprinted genes and loci has been identified in previous studies, genome-wide loss of imprinting in cancers has not been characterized fully to date.

Perturbation of imprinted gene expression also has been linked to chronic complex disorders. Higher expression of *IGF2* attributable to a loss of *IGF2* imprinting has been observed in rheumatoid arthritis patients (Martin-Trujillo et al., 2010), although whether this is a consequence or a cause of the immune response remains unclear. *IGF2* in peripheral blood is imprinted, indicating that *IGF2* is expressed monoallelically in unstimulated immune cells. Interestingly, cultured T cells exposed to phytohemagglutinin, an immune cell stimulant, showed 10 to 20-fold increased proliferation and exhibited loss of *IGF2* imprinting, which resulted in up to a six-fold increase in *IGF2* expression compared to unstimulated T cells (Hofmann, Takeuchi, Frantzen, Hoelzer, & Koeffler, 2002). In addition to genomic imprinting, reactivation of silenced X-linked alleles in females has been hypothesized to explain the high prevalence of autoimmune disorders in women (Libert, Dejager, & Pinheiro, 2010). For example, demethylation of *CD40L* on the inactive X chromosome has been observed in T cells of females with systemic lupus erythematous, a female-predominant autoimmune disease

(Lu et al., 2007).

4.4.3 Environmental effects

It is known well that environmental factors can influence gene expression through epigenetic modifications; for example, epigenetic ASEs, such as genomic imprinting, in which a particular allele is programmed to express or not, determine the dosage of gene expression. Thus, environmental factors could play a role in modulating ASEs. Bisphenol A (BPA) is a synthetic organic compound that is used widely in the production of plastics. BPA mimics estrogen properties and is an endocrine-disrupting agent. Increasing evidence has suggested that BPA has a wide range of effects on the brain and behavioral functions, including structural development of the brain, social behavior, anxiety and novelty (Wolstenholme, Rissman, & Connelly, 2011). Surprisingly, Bartolomei and colleague revealed that maternal BPA exposure during the early stages of embryonic development disrupted imprinting of several imprinted genes in placenta and embryos that are associated with changes in gene expression and alterations of DNA methylation in the differential methylated region of the imprinted loci (Susiarjo, Sasson, Mesaros, & Bartolomei, 2013). In addition to chemical compounds, diet, a well-studied environmental factor in gene expression, has been shown to regulate ASE. A methyl donor deficient diet can lead to the loss of *Igf2* imprinting in the mouse during the postweaning period (Waterland, Lin, Smith, & Jirtle, 2006). A recent study showed that protein restriction in utero activated the silenced paternal copy of *Cdkn1c*, an imprinted gene normally expressed maternally, in the embryo that persisted into adulthood even after the regular diet was resumed (Van de Pette et al., 2017). Both studies also

demonstrated that the loss of imprinting was associated with dysregulation of epigenetic marks. Although additional environmental factors also have been proposed to affect genomic imprinting, most of the implications relied upon indirect evidence (the change of gene expression and DNA methylation; Kappil, Lambertini, & Chen, 2015).

4.5 Potential mechanisms of scattered RME

The regulatory mechanisms of known RMEs have been explored in different studies in which some working models have been proposed (See Introduction). Apparently, one major difference between known and scattered RME genes is genomic location. Known RME genes, such as olfactory receptors and protocadherin, are clustered in the genome, unlike scattered RME genes. This suggests that the monoallelic expression of scattered RME genes is not attributable to a single common cis-acting regulatory factor. In addition, the regulatory mechanism must fulfill two criteria: first, it must choose the allele expressed randomly, and second, only one allele can be expressed. Based on these principles, two potential regulatory models of scattered RME genes were proposed. The first (Figure 4.1) was adapted from the enhancer competition model underlying stochastic choice of olfactory receptors (Lomvardas et al., 2006; Savarese & Grosschedl, 2006). Each mouse olfactory neuron expresses stochastically only one of the 1400 olfactory receptors (ORs) that cluster on different chromosomes. Lomvardas et al. found that H enhancer, located on chromosome 14, interacts with an OR gene either in cis (MOR28 on chromosome 14) or in trans (M50 on chromosome 7 or M71 on chromosome 9; Lomvardas et al., 2006). Thus, to be activated, ORs located on different chromosomes compete for a single enhancer. When one allele initiates a functional

expression, that expression produces negative feedback to prevent the activation of the H enhancer on the other allele. One known negative feedback mechanism is that translation of ORs triggers an unfolded protein response to downregulate *LSD1*, a histone demethylase; therefore, other silenced genes are not activated thereafter. Thus, these negative feedback mechanisms guarantee that only one allele can be expressed. This interchromosomal enhancer competition followed by negative feedback regulation may provide a potential mechanism for the way in which scattered RME genes express one allele randomly.

The dissertation's second hypothetical model is based on the dynamic RME model (Figure 4.2). Sandberg and colleague proposed that asynchrony of burst frequency and discrepancies in the rate of RNA degradation between two alleles lead to RME. Dynamic RME could account for the existence of monoallelic expression and the mixture of monoallelic and biallelic expressing cells in tissue samples. However, this model does not address the underlying epigenetic mechanisms of random transcriptional bursting. Here, a random enhancer binding is proposed that may explain this random bursting. In the hypothesized model, two genes compete randomly for one enhancer on each allele, which may lead to two different outcomes: monoallelic or biallelic expression. When one gene is activated on one allele and the other is activated on the other allele, both genes will show monoallelic expression. On the other hand, one gene will show biallelic expression and the other will be silenced when only one gene binds to the enhancer on both alleles. Thus, the change in bi/monoallelic expression may be determined by how many of the gene's alleles interact with the enhancer.

4.6 References

Angermueller, C., Clark, S. J., Lee, H. J., Macaulay, I. C., Teng, M. J., Hu, T. X., ... Kelsey, G. (2016). Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nature Methods*, *13*(3), 229-232.

Anwar, S. L., Krech, T., Hasemeier, B., Schipper, E., Schweitzer, N., Vogel, A., . . . Lehmann, U. (2012). Loss of imprinting and allelic switching at the DLK1-MEG3 locus in human hepatocellular carcinoma. *PLoS One, 7*(11), e49462.

Bonifazi, E., Gullotta, F., Vallo, L., Iraci, R., Nardone, A. M., Brunetti, E., ... Novelli, G. (2006). Use of RNA fluorescence in situ hybridization in the prenatal molecular diagnosis of myotonic dystrophy type I. *Clinical Chemistry*, *52*(2), 319-322.

Byun, H. M., Wong, H. L., Birnstein, E. A., Wolff, E. M., Liang, G., & Yang, A. S. (2007). Examination of IGF2 and H19 loss of imprinting in bladder cancer. *Cancer Research*, *67*(22), 10753-10758.

Cui, H. (2007). Loss of imprinting of IGF2 as an epigenetic marker for the risk of human cancer. *Disease Markers*, *23*(1-2), 105-112.

Cui, H., Cruz-Correa, M., Giardiello, F. M., Hutcheon, D. F., Kafonek, D. R., Brandenburg, S., . . . Feinberg, A. P. (2003). Loss of IGF2 imprinting: A potential marker of colorectal cancer risk. *Science, 299*(5613), 1753-1755.

Deng, Q., Ramskold, D., Reinius, B., & Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science, 343*(6167), 193-196.

Eckersley-Maslin, M. A., Thybert, D., Bergmann, J. H., Marioni, J. C., Flicek, P., & Spector, D. L. (2014). Random monoallelic gene expression increases upon embryonic stem cell differentiation. *Developmental Cell*, 28(4), 351-365.

Faddah, D. A., Wang, H., Cheng, A. W., Katz, Y., Buganim, Y., & Jaenisch, R. (2013). Single-cell analysis reveals that expression of nanog is biallelic and equally variable as that of other pluripotency factors in mouse ESCs. *Cell Stem Cell*, *13*(1), 23-29.

Filipczyk, A., Gkatzis, K., Fu, J., Hoppe, P. S., Lickert, H., Anastassiadis, K., & Schroeder, T. (2013). Biallelic expression of nanog protein in mouse embryonic stem cells. *Cell Stem Cell*, *13*(1), 12-13.

Gendrel, A. V., Attia, M., Chen, C. J., Diabangouaya, P., Servant, N., Barillot, E., & Heard, E. (2014). Developmental dynamics and disease potential of random monoallelic gene expression. *Developmental Cell*, *28*(4), 366-380.

Ginart, P., Kalish, J. M., Jiang, C. L., Alice, C. Y., Bartolomei, M. S., & Raj, A. (2016). H19 loss-of-imprinting mutant. *Genes & Development*, *30*(5), 567-578.

Guo, F., Li, L., Li, J., Wu, X., Hu, B., Zhu, P., ... Tang, F. (2017). Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells. *Cell Research*, *27*(8), 967.

Hansen, C. H., & Van Oudenaarden, A. (2013). Allele-specific detection of single mRNA molecules in situ. *Nature Methods*, *10*(9), 869-871.

Hofmann, W. K., Takeuchi, S., Frantzen, M. A., Hoelzer, D., & Koeffler, H. P. (2002). Loss of genomic imprinting of insulin-like growth factor 2 is strongly associated with cellular proliferation in normal hematopoietic cells. *Experimental Hematology*, *30*(4), 318-323.

Kappil, M., Lambertini, L., & Chen, J. (2015). Environmental influences on genomic imprinting. *Current Environmental Health Reports*, *2*(2), 155-162.

Kohda, M., Hoshiya, H., Katoh, M., Tanaka, I., Masuda, R., Takemura, T., ... Oshimura, M. (2001). Frequent loss of imprinting of IGF2 and MEST in lung adenocarcinoma. *Molecular Carcinogenesis*, *31*(4), 184-191.

Kriventseva, E. V., Koch, I., Apweiler, R., Vingron, M., Bork, P., Gelfand, M. S., & Sunyaev, S. (2003). Increase of functional diversity by alternative splicing. *Trends in Genetics*, *19*(3), 124-128.

Ku, C. J., Lim, K. C., Kalantry, S., Maillard, I., Engel, J. D., & Hosoya, T. (2015). A monoallelic-to-biallelic T-cell transcriptional switch regulates GATA3 abundance. *Genes & Development*, *29*(18), 1930-1941.

Larsson, C., Grundberg, I., Söderberg, O., & Nilsson, M. (2010). In situ detection and genotyping of individual mRNA molecules. *Nature Methods*, *7*(5), 395-397.

Lee, Y., & Rio, D. C. (2015). Mechanisms and regulation of alternative pre-mRNA splicing. *Annual Review of Biochemistry*, *84*, 291-323.

Levesque, M. J., Ginart, P., Wei, Y., & Raj, A. (2013). Visualizing SNVs to quantify allele-specific expression in single cells. *Nature Methods*, *10*(9), 865-867.

Libert, C., Dejager, L., & Pinheiro, I. (2010). The X chromosome in immune functions: When a chromosome makes the difference. *Nature Reviews Immunology*, *10*(8), 594-604.

Lomvardas, S., Barnea, G., Pisapia, D. J., Mendelsohn, M., Kirkland, J., & Axel, R. (2006). Interchromosomal interactions and olfactory receptor choice. *Cell, 126*(2), 403-413.

Lu, Q., Wu, A., Tesmer, L., Ray, D., Yousif, N., & Richardson, B. (2007). Demethylation of CD40LG on the inactive X in T cells from women with lupus. *The Journal of Immunology*, *179*(9), 6352-6358.

Manodoro, F., Marzec, J., Chaplin, T., Miraki-Moud, F., Moravcsik, E., Jovanovic, J. V., . . . Debernardi, S. (2014). Loss of imprinting at the 14q32 domain is associated with microRNA overexpression in acute promyelocytic leukemia. *Blood, 123*(13), 2066-2074.

Martin-Trujillo, A., van Rietschoten, J. G., Timmer, T. C., Rodríguez, F. M., Huizinga, T. W., Tak, P. P., ... Verweij, C. L. (2010). Loss of imprinting of IGF2 characterizes high IGF2 mRNA-expressing type of Fibroblast-like Synoviocytes in Rheumatoid Arthritis. *Annals of the Rheumatic Diseases, 69,* 1239-1242.

Medema, R. H., & Burgering, B. M. (2007). The X factor: Skewing X inactivation towards cancer. *Cell, 129*(7), 1253-1254.

Melamud, E., & Moult, J. (2009). Stochastic noise in splicing machinery. *Nucleic Acids Research*, *37*(14), 4873-4886.

Miyanari, Y., & Torres-Padilla, M. E. (2012). Control of ground-state pluripotency by allelic regulation of Nanog. *Nature, 483*(7390), 470-473.

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, *40*(12), 1413-1415.

Pedersen, I. S., Dervan, P. A., Broderick, D., Harrison, M., Miller, N., Delany, E., ... Tobin, B. (1999). Frequent loss of imprinting of PEG1/MEST in invasive breast cancer. *Cancer Research*, *59*(21), 5449-5451.

Perez, J. D., Rubinstein, N. D., Fernandez, D. E., Santoro, S. W., Needleman, L. A., Ho-Shing, O., . . . Dulac, C. (2015). Quantitative and functional interrogation of parent-of-origin allelic expression biases in the brain. *Elife, 4*, e07860.

Petropoulos, S., Edsgärd, D., Reinius, B., Deng, Q., Panula, S. P., Codeluppi, S., ... Lanner, F. (2016). Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell*, *165*(4), 1012-1026.

Pickrell, J. K., Pai, A. A., Gilad, Y., & Pritchard, J. K. (2010). Noisy splicing drives mRNA isoform diversity in human cells. *PLoS Genetics*, *6*(12), e1001236.

Reinius, B., & Sandberg, R. (2015). Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation. *Nature Reviews. Genetics*, *6*(11), 653.

Santoni, F. A., Stamoulis, G., Garieri, M., Falconnet, E., Ribaux, P., Borel, C., & Antonarakis, S. E. (2017). Detection of imprinted genes by single-cell allele-specific gene expression. *The American Journal of Human Genetics*, *100*(3), 444-453.

Savarese, F., & Grosschedl, R. (2006). Blurring cis and trans in gene regulation. *Cell, 126*(2), 248-250.

Steenman, M. J., Rainier, S., Dobry, C. J., Grundy, P., Horon, I. L., & Feinberg, A. P. (1994). Loss of imprinting of IGF2 is linked to reduced expression and abnormal methylation of H19 in Wilms' tumour. *Nature Genetics*, *7*(3), 433-439.

Stelzer, Y., Wu, H., Song, Y., Shivalila, C. S., Markoulaki, S., & Jaenisch, R. (2016). Parent-of-origin DNA methylation dynamics during mouse development. *Cell Reports*, *16*(12), 3167-3180.

Susiarjo, M., Sasson, I., Mesaros, C., & Bartolomei, M. S. (2013). Bisphenol a exposure disrupts genomic imprinting in the mouse. *PLoS Genetics*, *9*(4), e1003401.

Swanzey, E., & Stadtfeld, M. (2016). A reporter model to visualize imprinting stability at the Dlk1 locus during mouse development and in pluripotent cells. *Development, 143*(22), 4161-4166.

Van de Pette, M., Abbas, A., Feytout, A., McNamara, G., Bruno, L., To, W. K., . . . Fisher, A. G. (2017). Visualizing changes in Cdkn1c expression links early-life adversity to imprint mis-regulation in adults. *Cell Reports, 18*(5), 1090-1099.

Vorwerk, P., Wex, H., Bessert, C., Hohmann, B., Schmidt, U., & Mittler, U. (2003). Loss of imprinting of IGF-II gene in children with acute lymphoblastic leukemia. *Leukemia Research*, *27*(9), 807-812.

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., . . . Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature, 456*(7221), 470-476.

Waterland, R. A., Lin, J. R., Smith, C. A., & Jirtle, R. L. (2006). Post-weaning diet affects genomic imprinting at the insulin-like growth factor 2 (Igf2) locus. *Human Molecular Genetics*, *15*(5), 705-716.

Wolstenholme, J. T., Rissman, E. F., & Connelly, J. J. (2011). The role of Bisphenol A in shaping the brain, epigenome and behavior. *Hormones and Behavior*, *59*(3), 296-305.

Wu, H., Luo, J., Yu, H., Rattner, A., Mo, A., Wang, Y., . . . Nathans, J. (2014). Cellular resolution maps of X chromosome inactivation: Implications for neural development, function, and disease. *Neuron, 81*(1), 103-119.

Zuo, T., Wang, L., Morrison, C., Chang, X., Zhang, H., Li, W., . . . Liu, Y. (2007). FOXP3 is an X-linked breast cancer suppressor gene and an important repressor of the HER-2/ErbB2 oncogene. *Cell, 129*(7), 1275-1286.

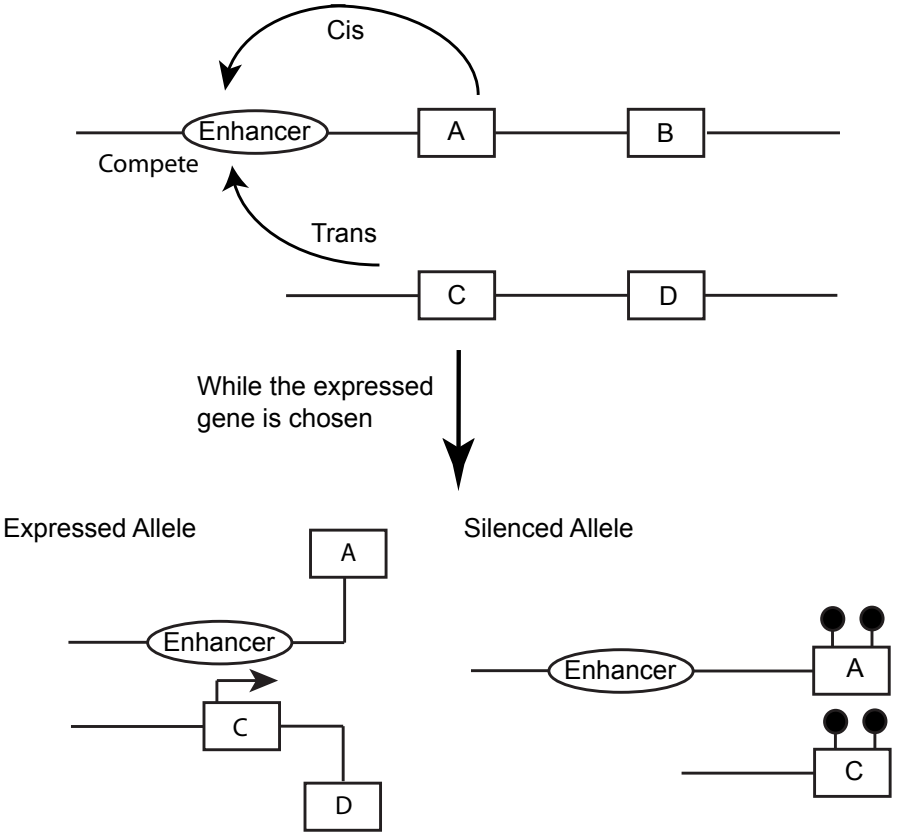


Figure 4.1 Enhancer competition model of RME. Several genes (A, B, C, D) compete for an enhancer in a cis or trans manner (Top). When one gene's allele starts a functional expression, that expression produces negative feedback to block the activation of other genes and of the other allele of the activated gene (Bottom).

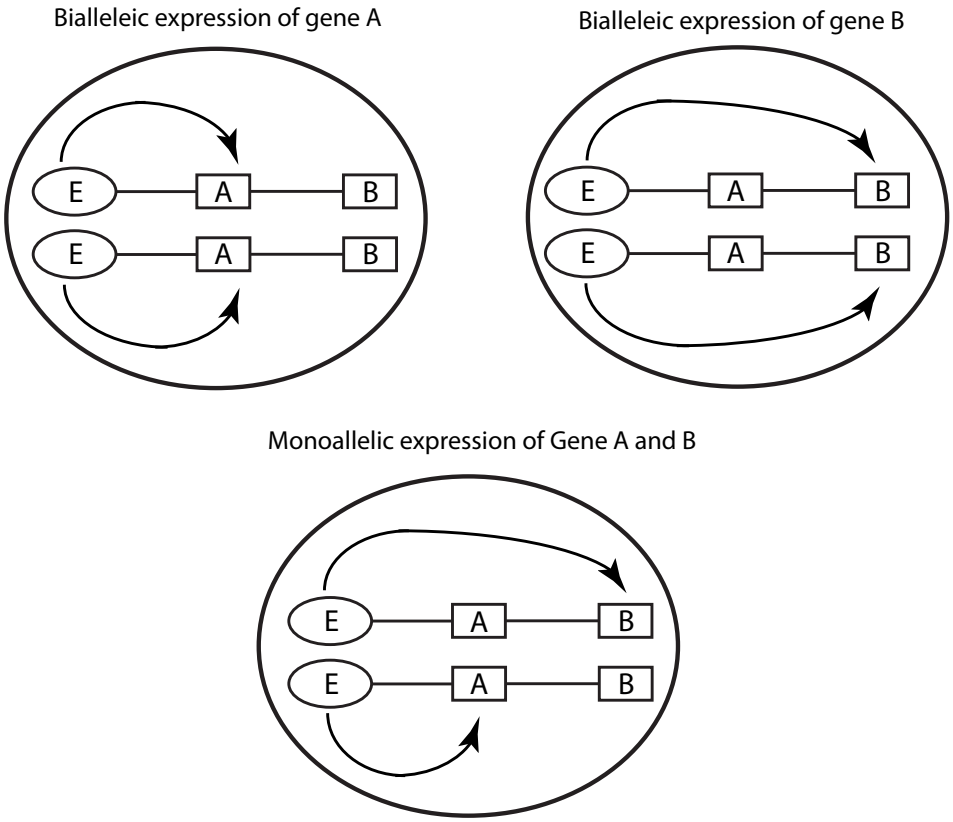


Figure 4.2 Random enhancer binding model of RME. Gene A shows biallelic expression and gene B is silenced when only gene A binds to the enhancer (E) on both alleles (Top left, the same mechanism for gene B at top right). When gene A is activated on one allele and gene B is activated on the other allele, both genes show monoallelic expression (Bottom).